Engineering the transition from non-living to living matter

Marc Rodríguez Garcia

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

School of Chemistry

College of Science and Engineering

August 2016

A tota la meva família, per haver-me ajudat a arribar fins aquí.

Però sobretot als meus pares, per estar tant a prop malgrat la distància.

I en especial a la Nuria, per ser el motiu que em fa tirar endavant.

"The good thing about science is that it's true whether or not you believe in it."

-Neil deGrasse Tyson

Acknowledgements

This project was carried out between September 2012 and June 2016 in the group of Prof Leroy Cronin in the School of Chemistry at the University of Glasgow. I have received much help and support from many colleagues and friends. In particular, I would like to express my sincere gratitude to the following people:

Prof Leroy Cronin, for taking me on as an Erasmus student, and then giving the opportunity to start a PhD in his research group. For his inspiration, guidance and constant support throughout these years.

Dr Andrew Surman, for his invaluable help and guidance during most of my PhD. Also, for teaching me to be rigorous and for teaching me everything I know about analysis.

Dr Geoff Cooper and Dr Soichiro Tsuda, for their kind help whenever it was needed**,** and for being such nice colleagues.

Dr Trevor Hinkley, for helping me during the first stages of my PhD.

Dr Mike Lee and **Zied Hosni**, for their help with LabView.

Kevin Donkers and **Dr Mohammed Hezwani,** for their contributions to the last part of this thesis.

Jonas Kube, **José Maria Planchat** and **Arielle le Nue**, for being amazing students and contributing so much to this work.

Dr Rebecca Turk Macleod and David Doran, for their help proofreading part of this thesis and for being such nice colleagues.

All the past and present members of the **Origins team**, for the nice discussions and the biscuits.

I would also like to thank the other PhD students who started with me: **Luzian Porwol**, **Qi Zheng**, **Juan Manuel Parrilla and James Taylor**, who helped make the writing process less of a tragic experience (and also **Dr Ross Winter).**

Sergio Martin, for being the best gym partner and a great friend.

Irene Suarez, for all her help for the past years and her friendship. I wish her the best in her PhD.

Mercè Martin, for the Mercè-facts, of which only 10% are true, but still enjoyable to hear.

Dr Laia-Vila Nadal, for being a great group coordinator and for her friendliness.

Stuart Marshall and **Amanda McGarvey,** for their help anytime I needed it.

Jim McIver, for his endless help setting all the platforms, the clean room… It's been a long way since I started in the attic!

Dr Diana Castro, for her technical support and kindness.

Also, **Dr Andreu Ruiz**, **Dr Ben Rausch**, **Dr Rachel Scullion**, **Dr Andrew Macdonell**, **Dr Jamie Cameron**, **Dr Salah Sharabi**, **Dr Ommid Anamimoghadam**, **Dr Christophe Busche**, **Dr Stefan Glatzel** for being great colleagues and friends.

All the members of the Cronin Group, past and present, who have made this such an enjoyable time, despite the horrendous weather.

Finally, **Dr Marc Segalés** and soon to be **Dr Irene Cascallana,** whom I shared this experience from the very first day I put my foot in Glasgow, and I wish them the best in their futures.

Table of Contents

Publications

The following articles were published as a result of work undertaken over the course of this PhD programme.

1. *Formation of oligopeptides in high yield under simple programmable conditions.* M. Rodriguez-Garcia, A.J. Surman, G. T. Cooper, I. Suarez-Marina, Z.Hosni, M. Lee and L. Cronin, *Nature Communications* **6,** (2015).

2. *A pathway to chemical evolution: environmental programming of peptide synthesis.* A.J. Surman, M. Rodriguez-Garcia, K. Donkers, J.M. Planchat, J. Kube, M. Hezwani and L. Cronin (In preparation).

3. *Scalable separation using ensembles and repetitions.* A. Henson, R. Turk, M. Rodriguez-Garcia and L.Cronin (In preparation).

4. *Microfluidic system for on-chip chemical evolution.* M. Rodriguez-Garcia and L.Cronin (In preparation).

Abbreviations

In addition to standard notation, the following abbreviations were used in this thesis:

Ala, A	Alanine
Asp, D	Aspartic Acid
BPC	Base-peak chromatogram
CAAs	Carbamoyl amino acids
CDI	Carbonyldiimidazole
COS	Carbonyl sulfide
DNA	Deoxyribonucleic acid
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ESI-MS	Electrospray Ionisation Mass Spectrometry
Glu, E	Glutamic Acid
Gly, G	Glycine
GlyGly, GG	Glycylglycine
His, H	Histidine
HPLC	High-Performance Liquid Chromatography
IP	Ion-pairing
IR	Infrared
LC	Liquid Chromatography
LC-MS	Liquid Chromatography-Mass spectrometry
Leu, L	Leucine
Lys, K	Lysine
MeCN	Acetonitrile
NCA	N-carboxy anhhydride
NMR	Nuclear Magnetic Resonance Spectroscopy

RNA	Ribonucleic acid
RP	Reverse Phase
PCA	Principal Component Analysis
PDMS	Polydimethyilsiloxane
Ser, S	Serine
SIPF	(Salt-Induced Peptide Formation)
Thr, T	Threonine
UV	Ultra-Violet
Val, V	Valine
vBPC	Virtual Base-peak chromatogram

Abstract

Re-creating and understanding the origin of life represents one of the major challenges facing the scientific community. We will never know exactly how life started on planet Earth, however, we can reconstruct the most likely chemical pathways that could have contributed to the formation of the first living systems. Traditionally, prebiotic chemistry has investigated the formation of modern life's precursors and their self-organisation under very specific conditions thought to be 'plausible'. So far, this approach has failed to produce a living system from the bottom-up.

In the work presented herein, two different approaches are employed to explore the transition from inanimate to living matter.

The development of microfluidic technology during the last decades has changed the way traditional chemical and biological experiments are performed. Microfluidics allows the handling of low volumes of reagents with very precise control. The use of micro-droplets generated within microfluidic devices is of particular interest to the field of Origins of Life and Artificial Life. Whilst many efforts have been made aiming to construct cell-like compartments from modern biological constituents, these are usually very difficult to handle. However, microdroplets can be easily generated and manipulated at kHz rates, making it suitable for high-throughput experimentation and analysis of compartmentalised chemical reactions. Therefore, we decided to develop a microfluidic device capable of manipulating microdroplets in such a way that they could be efficiently mixed, split and sorted within iterative cycles. Since no microfluidic technology had been developed before in the Cronin Group, the first chapter of this thesis describes the soft lithographic methods and techniques developed to fabricate microfluidic devices. Also, special attention is placed on the generation of water-in-oil microdroplets, and the subsequent modules required for the manipulation of the droplets such as: droplet fusers, splitters, sorters and single/multi-layer micromechanical valves.

Whilst the first part of this thesis describes the development of a microfluidic platform to assist chemical evolution, finding a compatible set of chemical building blocks capable of reacting to form complex molecules with endowed replicating or catalytic activity was challenging.

Hence, the second part of this thesis focuses on potential chemistry that will ultimately possess the properties mentioned above. A special focus is placed on the formation of peptide bonds from unactivated amino acids, despite being one of the greatest challenges in prebiotic chemistry. As opposed to classic prebiotic experiments, in which a specific set of conditions is studied to fit a particular hypothesis, we took a different approach: we explored the effects of several parameters at once on a model polymerisation reaction, without constraints on hypotheses on the nature of optimum conditions or plausibility. This was facilitated by development of a new high-throughput automated platform, allowing the exploration of a much larger number of parameters. This led us to discover that peptide bond formation is less challenging than previously imagined. Having established the right set of conditions under which peptide bond formation was enhanced, we then explored the co-oligomerisation between different amino acids, aiming for the formation of heteropeptides with different structure or function. Finally, we studied the effect of various environmental conditions (rate of evaporation, presence of salts or minerals) in the final product distribution of our oligomeric products.

1 *Introduction*

1.1 **The origin of life: an unresolved problem**

The transition from an inanimate inorganic world, principally consisting of minerals and gasses, to the formation of a minimal living system remains one of the greatest mysteries in science. It constitutes a particularly difficult subject to study because we do not have direct evidence about the conditions or the chemicals that were available on the early Earth. For this reason, we will never know exactly how life started on planet Earth. In this sense, our job as chemists is to reconstruct the most likely chemical pathways that could have contributed to the formation of the first living systems. As stated by Albert Eschenmoser: "Biogenesis, as a problem of science, is lastly going to be a problem of synthesis. The origin of life cannot be 'discovered', it has to be re-invented."

According to Alexander Oparin's and John Haldane's views (two pioneers in proposing a chemical origin of life), the first minimal living-systems on Earth originated via a series of chemical steps of increasing molecular complexity and functionality. It is now is widely accepted that this was a continuous process which took place between the late Hadean and the early Archean eon (4.1 to 3.5 billion of years ago) (Fig.1). This is corroborated by the discovery of potential biogenic carbon in 4.1 billion year old zircons[1] and the earliest potential chemofossils at 3.83 billions of years ago.[2]

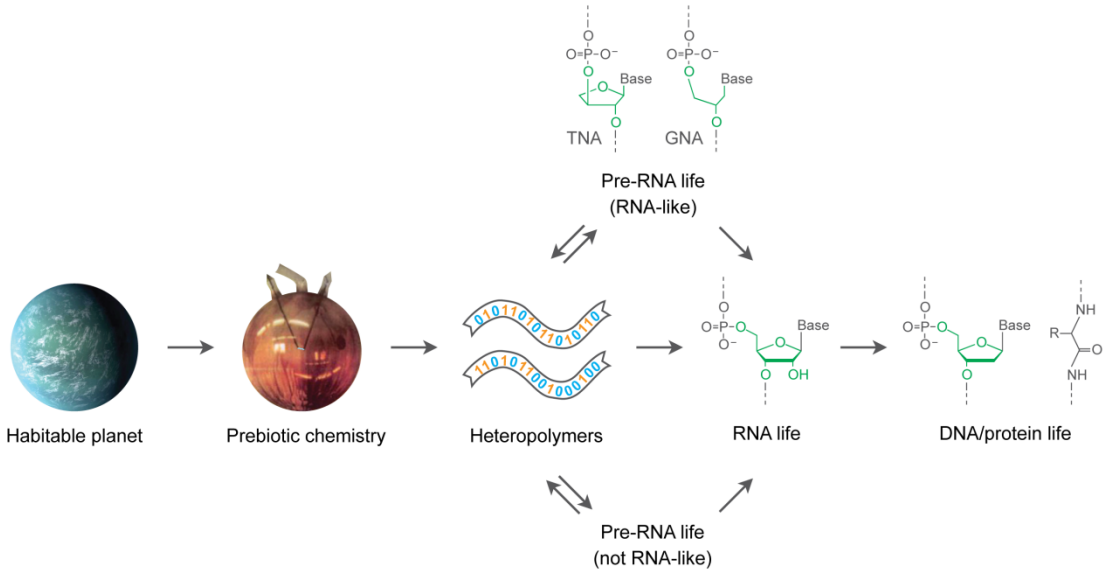


Figure 1: Chronological diagram showing the transition from prebiotic chemistry to the emergence of the first living organisms. Reproduced from Joyce, G. F.[3]

It is estimated that the total amount of extra-terrestrial organic matter delivered to our planet during the late Hadean and the early Archaean (known as the late heavy bombardment) was on the order of 10^{16}-10^{18} kg.[4] We are not certain whether the source of organic matter that progressively reacted and self-organised to give rise to the first living forms came from endogenous synthesis or exogenous delivery. However, during the last 50 years, we have started to find out about the chemistry that occurs in the space, the organic compounds recovered in meteorites and the possible geochemical scenarios that were present when life emerged in our planet. These findings facilitate to reduce the chemical inventory and the geochemical conditions from a vast parameter space, but at the same time, it has caused scientists to propose hypotheses on the origin of life under very constrained conditions, which usually results in their inability to demonstrate their hypothesis experimentally. But before we try to disseminate the different hypotheses on the origin of life, it is worth examining the greatest evidence of all: life itself. Despite modern life being the product of thousands of millions of years of evolution, and the evident differences between the vast number of species that live and have lived on this planet, there is a unifying mechanism that operates within all living beings: the Central Dogma of Molecular Biology. The Central Dogma, which was first proposed by Francis Crick, describes the flow of genetic information that operates within almost all biological systems, or as it is usually simplistically described: "DNA makes RNA makes Protein".

DNA and RNA are nucleic acid biopolymers made up of nucleotide monomers. Each nucleotide has three components: a nucleobase, a five-membered ring sugar and a phosphate group. There are two main structural differences between DNA and RNA: first, DNA uses the pyrimidine base thymine, whilst in RNA thymine is substituted by uracil. Secondly, DNA lacks a hydroxyl group in the 2' carbon in the ribose moiety, which makes it less susceptible than RNA to get damaged by hydrolysis. This feature allows a single strand DNA chain to form a double helix with a complementary chain. The role of DNA is to store genetic information, whilst RNA is used to translate the encoded information in DNA into protein biosynthesis. First, in a process called transcription, a single strand of DNA is copied into a complementary strand of RNA through an enzymatic process. The newly formed RNA strand (known as messenger RNA or mRNA) is then used by the ribosome to translate a series of three-nucleotide blocks (known as codons) into individual amino acids, which are subsequently linked to make a polypeptide chain. Therefore, there is a direct relationship between the final sequence of amino acids within the polypeptide chain and the original DNA molecule triplet codons.

The role of peptides and proteins is to provide functions such as transcription, regulation, transport, storage and structural support in the cells. This general process in which genetic information is translated into functional molecules is considered to be universal since it is found in almost all life on Earth.

Previously, it was thought that the genetic code was identical in all biological systems, and incapable of further evolution.[5] It was proposed that 61 different codons were specifically encoding for the 20 proteinogenic amino acids, and the remaining 3 were used as termination codons.[6]

However, in 1979, a variation in the established genetic code was first found in mitochondria.[7] Since then, other variants have been found, not only in mitochondrial but also in nuclear systems. This demonstrated that the genetic code is subject to change and capable of evolving.[8]

Also, the existence of three proteinogenic amino acids, other than the 20 standard ones, have been discovered: selenocysteine, pyrrolysine and N-formylmethionine.[9]

Nonetheless, the overwhelming universality in the basic mechanisms of transmitting genetic information and translating it into protein machinery across species suggests that all living beings evolved from a common ancestor, commonly known as LUCA (Last Universal Common Ancestor). This concept has been further corroborated by comparing ribosomal RNA (rRNA) sequences from different species, in which a high degree of correlation has been found. The modern phylogenetic tree of life (Fig. 2) has been built by comparing and relating these rRNA sequences, being the Last Universal Common Ancestor at the root of the tree. It has been suggested that LUCA was a '*simpler, more rudimentary entity than the individual ancestors that spawned the three groups (archaea, bacteria and eucharia)'*.[10] However, it has been pointed out that the even the last ancestor would have had a very complex metabolic machinery, with the ribosome being at the core of early protein biosynthesis.[11]

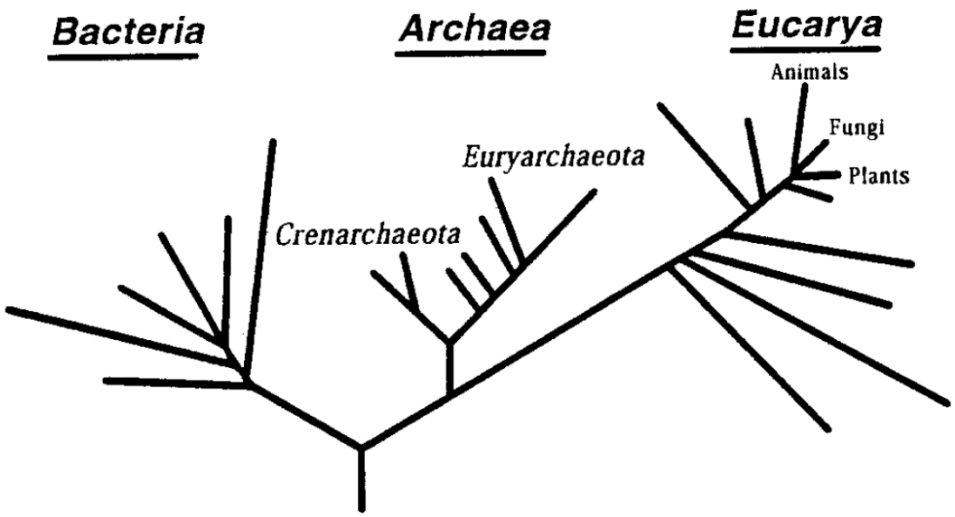


Figure 2: A simple representation of the tree of life, encompassing the diversity represented by sequenced genomes. Adapted with permission from National Academy of Sciences. Copyright (2000) National Academy of Sciences.[12]

At this point, considering that the earliest life forms would already be highly evolved, we need to understand the transition by which simple organic molecules increasingly become more complex to finally give rise to the simplest unit of life. Given that we have no knowledge about the nature of the proto-life forms that existed before Last Universal Common Ancestor, it is important to have a definition of life that help us to delineate which are the essential properties that a chemical system requires to be considered alive. In that process, exploring the transition from inanimate to living matter is not only going to help us to propose a solution to the origin of life, but it will also help us to elucidate the most fundamental properties of biological systems. Furthermore, we might be able to create living systems in the laboratory that are not based in the same chemistry.

Despite many different definitions, and in order to distinguish a non-living from a living system, it is widely accepted that a living system must possess the following properties:

- The ability to encapsulate its components in an enclosed boundary and distinguish itself from the environment.
- The ability to transmit information from one generation to the next.
- The ability to accrue variations (mutations) in the attributes between successive generations and undergo Darwinian Evolution based on natural selection.

However, a simpler and generally accepted definition was proposed by NASA: life is "a self-sustaining chemical system capable of Darwinian evolution."[13]

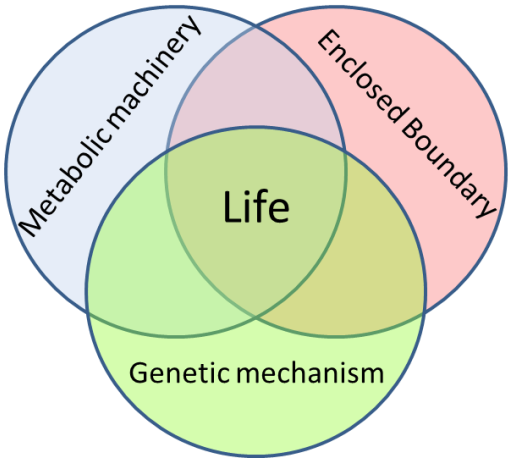


Figure 3: Life consists of three main interdependent components.

Whilst it is generally accepted that life is based in three main interdependent components (Fig. 3), different approaches have been explored to construct a minimal living system. The classic approach explores the formation of the three components of life from an Origin of Life point of view by attempting to synthesise the precursors and molecules that constitute modern life forms under the conditions thought to be present on the early Earth. On the other hand, the field of Artificial Life attempts the formation a chemical system that possesses the minimal requirements to be considered alive, with no constraints on the chemical nature of the components in the system.

Whilst Artificial Life is a relatively new and exciting field of study, most of the work performed during the second half of the 20th century aimed to create a living system based on the chemistry of modern cells. Usually, these experiments were performed to fit a particular hypothesis which was usually restricted by very specific conditions on hypothetical geochemical scenarios or specific starting materials.

In this sense, many theories about the Origin of Life have been postulated but none of them has been able to produce life under laboratory conditions.

1.1.1 Historical Perspective

Through the course of modern human history, it was generally assumed by naturalists and philosophers that spontaneous generation was the most plausible explanation for the origin of life. It was not until the 19th century when this idea started to be challenged.

A few early examples on the synthesis of organic compounds from inorganic starting materials played a key role in understanding the transition from non-living to living matter. The first example should be attributed to Friedrich Wohler, who in 1828 achieved the synthesis of urea from ammonium cyanate. Further examples such as the Strecker synthesis in 1850 (synthesis of alanine from a mixture of acetaldehyde, ammonia and hydrogen cyanide) and the Formose reaction in 1861 by Butlerov (sugars synthesis from formaldehyde in strong alkaline conditions) represent the earliest work on the abiotic synthesis on some of life's building blocks, although the researchers who performed these studies were not interested in how life began on Earth.

Despite the significance of these discoveries, there still was a great support to spontaneous generation as an explanation for the Origin of Life. However, the idea that living organisms were the result of a gradual transformation process of inanimate matter started to become popular after the publication of *The Origin of Species* in 1859. Darwin did not include in his book specific details on his views on a chemical Origin of Life rather than mentioning that *"all organic beings which have ever lived on this Earth have descended from one primordial form, into which life was first breathed".* Twelve years after the publication of his books, Darwin sent a letter to Joseph Dalton Hooker in which he had the idea that life could have emerged in a warm little pond containing all sorts of ammonia, phosphoric acids, light, heat and electricity.

Inspired by Darwin, Pasteur then demonstrated the implausibility of spontaneous generation after showing that microorganisms could not grow in sterile sealed flasks in 1859. For a few decades, there was a general belief that living organisms were the product of a gradual transformation from inanimate matter. However, most of the hypotheses proposed during those years were incomplete or went unnoticed. Other scientists, including Lord Kelvin (one of University of Glasgow's greatest scientists) suggested alternative theories that involved the delivery of microbes from other planets and an extra-terrestrial origin of life. These theories are known as Panspermia.

It was not until the 1920s when two scientists, Alexander Oparin and John Haldane independently revisited the idea that life on Earth originated via a series of chemical steps of increasing molecular complexity and functionality. They suggested that in a highly reducing atmosphere, organic compounds would have formed. Then, aggregation of these organic compounds led to the formation of coacervates, from which the first heterotrophic microbes evolved. This hypothesis has since been known as the 'Primordial Soup theory'.

1.1.2 The birth of Prebiotic Chemistry

Oparin's most famous book, called *Origin of Life* (1936) went unnoticed for a few years until it caught the attention of biologists during the 1950s. Preliminary experiments performed by Melvin Calvin attempted the synthesis of organic compounds under primitive Earth conditions.[14]

It was not until 1952 when Harold Urey, a chemistry professor at the University of Chicago, was approached by a postgraduate called Stanley Miller, who proposed to run an experiment that could produce the essential organic compounds that were thought to be relevant for the origin of life on Earth. Their initial aim was to design an experiment that collected the components that could have contributed the development of chemical reactions on the early earth such as: the presence of atmosphere, an ocean and energy. They then designed and built an apparatus comprised of a round-bottom flask connected to a bigger round-bottom flask with a set of electrodes that generated high-voltage electrical sparks. One of the flasks was filled with water to simulate the ocean, which was heated up until evaporation. The gas water molecules were then transferred to a bigger flask that contained a mixture of gases that emulated the early earth's atmosphere (hydrogen, methane and ammonia), which were continuously passed through electrical sparks to simulate lightning. The reacted species in the vapour phase were then condensed and re-circulated to the water-containing flask.

It only took a year to publish the results from what is considered to be one of the most important experiments from the 20th century.[15] The first results showed the appearance of amino acids such as glycine, alanine, β -alanine, α-aminobutyric acid and aspartic acid. A few years later, Miller discovered that mechanism by which the formation of amino acids occurred in his famous apparatus was via the Strecker synthesis.[16]

Hydrogen cyanide and aldehydes were formed when the gas mixture was exposed to spark discharges. These would then react in water and in the presence of ammonia to form amino acids.

This work was the first to demonstrate that a variety of the biochemically significant compounds could be synthesised from simple molecules, proving Oparin's hypothesis. These astonishing results established the origin of life research and prebiotic chemistry as emerging fields of research.

A few years after Miller's famous publication in Science, a young Catalan biochemist managed to achieve another breakthrough: the synthesis of the nucleobase adenine under laboratory conditions. Joan Oró achieved the synthesis of adenine by reacting hydrocyanic acid in the presence of ammonia. The results were published in *Nature*[17], where he described the mechanism by which 5 molecules of HCN condensed into 1 molecule of adenine. A very interesting observation from these results was that adenine (an essential constituent of nucleotides) could be synthesised from HCN (a lethal compound).

Just after the publication of his extraordinary result, Oró also achieved the synthesis of various amino acids (glycine, alanine and aspartic Acid) from the same starting materials.[18]

Later in his career, Oró suggested that cometary collisions could have contributed to increasing the amount of organic matter that could have resulted in the synthesis of biochemical molecules on the early earth.[19]

The abiotic synthesis of biochemical monomers, such as amino acids and purines was achieved within 10 years. After that, many scientists became interested in an experimental approach to study the Origin of Life, which later resulted in the birth of prebiotic chemistry as a field of study. The discoveries from Miller and Oró, as well as the later detection of relevant biochemical molecules in meteorites marked a real turning point on the study and understanding on the Origin of Life.

It also made evident that the monomers of relevant biochemical polymers (RNA and proteins) and the constituents of membranes could have been available from many sources (endogenous synthesis or exogenous delivery).

1.1.3 Main Theories on the Origin of Life

The main function of DNA is to store genetic information, and it has no functional or catalytic properties. On the contrary, proteins cannot store genetic information but they are responsible for regulating the metabolic processes within the cell. Given that DNA requires highly complex and specific proteins to support its replication, and proteins are synthesised according to the information stored in DNA, how could these different molecules have arisen without each other? This conundrum is usually referred to as the 'Egg-chicken' paradox.

Given the progress during the 1950s on the abiotic synthesis of some of life's precursors, the idea of a heterotrophic origin of life became generally accepted. A huge effort was then placed on the synthesis of the building blocks of life under prebiotic conditions and their further polymerisation. Based on the initial success of these experiments and the development of new techniques that helped to elucidate the structure of both nucleic acid and proteins, several theories on the Origin of Life were proposed, which are mainly divided in the 'Replication-first' or the 'Metabolism-first' theories.

1.1.3.1 Replication-first hypothesis

After the discovery of DNA's double helix structure, it was also discovered that single-stranded RNA chains could form well-defined tertiary structures. This led Carl Woese, Leslie Orgel and Francis Crick to propose the idea that the first living entities used RNA both to store genetic information and to perform metabolic functions.

This theory did not have a lot of support until the 1980s, when Cech and Altman independently discovered that short RNA molecules could act as enzymes[20,21]. Since then, this type of RNA molecules became known as ribozymes. Ever since, the so called 'RNA world' theory was established as the most likely explanation for the Origin of Life on Earth.[22–24]

Later discoveries contributed to the consolidation of this theory: ribosomes, which are central to protein biosynthesis and thought to be one of the most conserved features in the earliest life forms, are mainly composed of RNA. Also, transfer RNAs are essential for amino acid activation in protein biosynthesis.

Furthermore, ATP is used as the energy carrier in the cells. All this evidence suggests that RNA had a key role in the transition towards the first life forms.

Despite being one of the most accepted theories, the 'RNA world' faces fundamental problems that still remain to be solved. For the last 50 years, many scientists have tried to synthesise RNA from a bottom-up approach. Attempting to synthesise their monomeric building blocks has not yet been entirely solved, and many arguments about the plausibility of these reaction center the debate amongst researchers. A thorough analysis of the state-of-the art on the abiotic synthesis of RNA is described in section 1.3.2.

1.1.3.2 Metabolism-first hypothesis

Metabolism-first hypotheses agree on the idea that chemical networks with a high degree of mutual catalysis between its components allow for adaptation and evolution without any molecular replication.

One of the most relevant 'Metabolism-first' hypotheses was proposed by Wächtershäuser in 1988 under the name of the 'Iron-Sulfur (Fe-S) World Hypothesis'.[25]

As opposed to the 'Primordial Soup' theory, in which the main building blocks of life are synthesised with the aid of external sources of energy (UV radiation, lightning), the Fe-S hypothesis proposes that the synthesis of organic compounds was driven by the energy released from redox reactions of metal sulphides. The ideal scenario where these processes could have taken place is in deep-sea hydrothermal vents. It was later discovered that iron sulphide mineral catalysed a series of chemical reactions that create a reverse citric acid cycle, in which carbon monoxide is reduced to form complex organic molecules that are central in modern life forms metabolism, such as acetate and pyruvate.[26] It has also been shown that amino acids and dipeptides can result from this reaction.[27,28] Apart from producing relevant organic molecules, it was observed that Fe-S complexes create microscopic bubbles upon precipitation, which has led to the hypothesis that the first cells were not using lipid membranes but these capsules instead[29].

Whilst this hypothesis has many associated problems (for example, it does not explain how these metabolic cycles could reproduce or evolve in the absence of genetic material), it provided an alternative framework for understanding the origin of life in the absence of highly complex molecules like RNA.

A significant amount of research has focused on the reactions that occur in hydrothermal systems, not only for the development of proto-metabolic cycles, but also for understanding the missing-link between metabolism-first scenarios and replication-based life forms.[29–31]

1.1.3.3 Alternative theories

Since the consolidation of the 'RNA-world' as one the most plausible hypothesis to explain the emergence of life, alternative hypotheses suggested that peptides and proteins played a major role, as opposed to an RNA-only model.

'Peptide-first' scenarios, such as the GADV-Protein World hypothesis[32] suggest that the emergence of auto-catalytic networks of self-reproducing proteins preceded the first RNA-based proto-cells.[33] These proteins would only consist of four amino acids (glycine, alanine, aspartic acid and valine), and their accumulation due to pseudo-replication processes would drive the formation of oligonucleotides enzymatically.[34] However, the nature of these GADV proteins resembles the *Proteinoids* described by Fox[35], in which the fraction of peptide bonds is very low.

Ikehara proposed that the most primitive genetic code would only use 3 codons (GNC genetic code) to code for 4 amino acids.[32] Nonetheless, Ikehara's work does not describe how the structure and functionality of the proposed GNC codons could have occurred.

A 'Lipid world' theory has also been proposed as an early evolutionary step in the emergence of cellular life. This theory postulates that lipid –like objects were able to self-replicate and transmit composomal information to their progenies.[36]

Whilst there is still a lot of debate between RNA-first or metabolism-first scenarios, there is still no sufficient evidence supporting one theory or the other. RNA-first supporters still face the complicated abiotic synthesis of nucleotides, whilst the inability to form long peptides disproves a metabolism-first scenario.

However, some research on the Origin of Life is starting to divert into scenarios where no preferred molecules are chosen, given the limitations on the bottom-up synthesis of both RNA and proteins. These scenarios propose that life arose from an early symbiosis between amino acid and nucleotide chemistries (Fig. 4).[37]

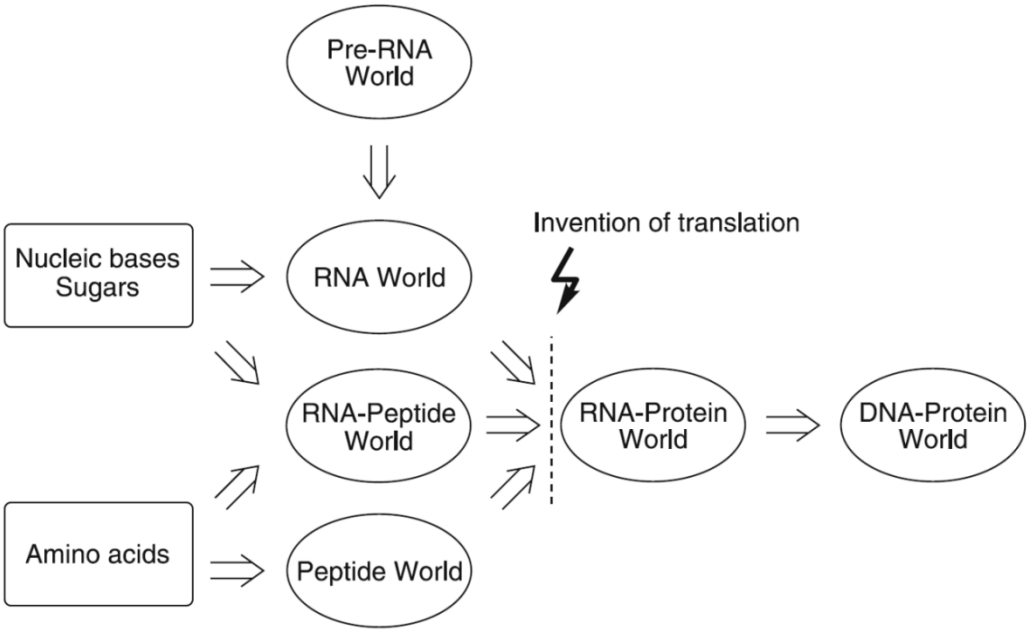


Figure 4: The sequence of different hypothetical stages in the evolution from an abiotic chemical world towards the modern DNA-protein world. Reprinted with permission from Springer, Copyright © 2005, Springer.[38]

The next section summarises the work that has focused on the synthesis of relevant biochemical monomers and the problems associated with some of them for the last 60 years.

1.2 Synthesis of monomers

1.2.1 Synthesis of Amino acids

As described in the previous section, the first experiment that demonstrated the formation of various amino acids under conditions that were aimed to simulate a prebiotic environment was the Miller-Urey experiment. However, the first synthesis of an amino acid under abiotic conditions should be attributed to Adolph Strecker, when in 1850 managed to synthesise alanine from acetaldehyde, ammonia and hydrogen cyanide.[39]

As stated later by Miller, the mechanism by which amino acids were produced in his spark discharge experiment was the Strecker synthesis. He proposed that aldehydes and hydrogen cyanide were synthesised in the gas phase by exposure to the spark. Then, these would further react in the aqueous solution in the presence of ammonia to produce α-

aminonitriles and cyanohydrins. Slow hydrolysis of these products would finally yield a mixture of α-amino acids and α-hydroxy acids.[16]

Since then, many other researchers tried to reproduce Miller's original work by changing various parameters, in order to obtain different mixtures that could help to elucidate which molecules could have been present on the early earth (Fig. 5).[40] Also, whilst the initial choice of gases suggested that a reducing atmosphere was required to produce amino acids, later studies showed that under neutral conditions, amino acid production was still efficient.

Authors	Reactants	Energy source	Results reported
Miller [5]	CH_4, NH_3, H_2O, H_2	Electric discharges	Simple amino acids, organic compounds
Abelson [6]	CO, CO_2, N_2, NH_3, H_2, H_2O	Electric discharges	Simple amino acids, HCN
Bar-Nun et al. [7]	CH_4, NH_3, H_2O	Shock wave	Simple amino acids
Harada and *Fox* [8]	CH_4, NH_3, H_2O	Thermal energy (900–1200°)	14 proteinogenic amino acids
Lawless and *Boynton* [9]	CH_4, NH_3, H_2O	Thermal energy	Glycine, alanine, aspartic acid, β-alanine, *N*-methyl-β-alanine, β-aminobutanoic acid
Groth and *Weyssenhoff* [10]	CH_4, NH_3, H_2O	Ultraviolet light (1470 and 1294 Å)	Simple amino acids (low yields)
Sagan and *Khare* [11]	CH_4, C_2H_6, NH_3, H_2O, H_2S	Ultraviolet light (> 2000 Å)	Simple amino acids (low yields)
Yoshino et al. [12]	H_2, CO, NH_3, montmorillonite	Temperature of 700°	Glycine, alanine, glutamic acid, aspartic acid, histidine, lysine, arginine
Palm and *Calvin* [13]	H_2, CH_4, NH_3, H_2O	Electron irradiation	Glycine, alanine, aspartic acid
Miyakawa et al. [14]	CO, N_2, H_2, H_2O	High-temperature plasma	Glycine, alanine, aspartic acid
Rode and co-workers [38]	CO_2, N_2, H_2O	Electric discharges	Glycine, alanine, valine, serine, proline, lysine, histidine

Figure 5: Some Examples of Miller-Type Experiments Simulating Assumed Primordial Earth Scenarios. Adapted with permission from John Wiley and Sons, Copyright © 2007 Verlag Helvetica Chimica Acta AG, Zürich.[41]

However, it was not until 2008 when a team of researchers lead by Jeffrey Bada (a former student of Miller's) re-analysed some of the 1950s Miller-Urey experiment samples that were kept in an old box.[42] The use of modern analytical techniques allowed the detection of up to 22 amino acids, as well as hydroxylated amino acids and amines, in contrast to the 7 initial amino acids reported by Miller in his 1953 publication in *Science*. In 2014, Parker *et al.* re-analysed and reproduced a never-reported Miller experiment which included cyanamide (a prebiotic condensing reagent) and reported the formation of a variety of amino acids, glycine dipeptides and diketopiperazines.[43]

As evidenced by an extensive amount of work, the Strecker synthesis represents the simplest pathway by which amino acids can be synthesised from prebiotically available compounds. Alternative pathways, such as the Bücherer-Bergs reaction, in which the addition of CO_2 promotes the hydrolysis of aminonitriles produced in the Strecker synthesis, can yield hydantoins (amino acid precursors).

One of the intermediate products obtained in the formation of hydantoins are carbamoyl amino acids, which have been recently considered as a form of activated amino acid monomers for the synthesis of polypeptides.[44]

Whilst many experiments were carried out to find out conditions under which amino acids could have been synthesised endogenously, the discovery of a rich-variety of organic compounds in carbonaceous chondrites (including amino acids) opened the possibility for an alternative source of amino acids and other relevant biochemical compounds.

Carbonaceous chondrites are a very rare type of meteorites since very few of them have been collected since the 19th century (they represent just 4.7% of the total collected), and are one of the few types of meteorites that contain organic carbon. The most famous carbonaceous chondrite reported is the Murchison meteorite (named after the small town in Australia where it fell). The Murchison meteorite has become one of the most well characterised meteorites since it is the most organic-rich meteorite known to have fallen on earth. The majority of the organic carbon found in carbonaceous chondrites corresponds to an insoluble macromolecular material. However, a small fraction of organic compounds are found in very low concentration (0-100ppm).[45] These represent a highly complex mixture containing hydrocarbons, carboxylic acids, amino acids, hydroxy acids, amines, amides, etc. Whilst more than 70 amino acids have been detected, only 8 of them were proteogenic.[46]

The distribution of molecular structures (decreasing abundance with the number of carbon atoms and predominance of branched isomers) suggests that the mechanism of formation of these compounds could be associated to a non-selective photochemical or radical process.

Also, the possibility that amino acids could be synthesised in an interstellar medium has been discussed. Since the detection of glycine molecules in interstellar dense molecular clouds[47], a number of experiments have been performed to reproduce and understand the interstellar medium reactions. Several studies have reported the formation of amino acids when interstellar ice analogues (mixtures of H_2O, NH_3, CO, CO_2, MeOH) were exposed to UV or bombarded high-energy protons.[48–51] Also, atmospheric shock impact synthesis could have been an important source of terrestrial organics. In a recent report by Martins *et al.*[52], the formation of several α-amino acids was achieved when an icy mixture containing NH_4OH, CO_2 and CH_3OH was shocked with a steel projectile fired at a very high velocity.

1.2.2 Synthesis of Nucleotides

The difference between amino acids and nucleotides as monomeric units of biopolymers is that nucleotides, as opposed to amino acids, are composed of three sub-monomeric parts: a nucleobase (adenine, guanine, cytosine, and uracil in RNA or thymine instead of uracil in DNA), a sugar (ribose or deoxyribose in DNA), and a phosphate group. As described in the previous section, the synthesis of amino acids under abiotic conditions proceeds readily from very simple compounds. However, the bottom-up synthesis of nucleotides has represented one of the major synthetic challenges of the last half of the 20th century, and still remains partially unsolved. Historically, since nucleotides are composed of three sub monomeric parts, prebiotic chemists have tried to independently synthesise those parts to couple them afterwards. Whilst this approach has not succeeded to produce nucleotides, alternative pathways have been recently explored. The next section aims to describe the most relevant work performed on the abiotic synthesis of nucleotides and the problems associated with it.

After Joan Oró reported the synthesis of adenine from HCN in the presence of ammonia in 1961, following-up experiments were performed to achieve the synthesis of the other canonical nucleobases.

For example, the synthesis of pyrimidine nucleosides was reported by Orgel, when cyanoacetylene (a product abiotically obtained from the electric discharge of a mixture of methane and nitrogen) was reacted with simple inorganic substances such as urea.[53] Nucleobases have also been obtained by using formamide as a starting material under different conditions: by UV irradiation, in the presence of inorganic catalysts and in the presence of minerals.[54] Whilst all the aforementioned examples describe the formation of canonical nucleobases, the products of these reactions are often very complex, and contain a variety of other nucleobase derivatives.

1.2.2.1 Synthesis of sugars

The synthesis of ribose (the sugar moiety in nucleotides) is more troublesome than the synthesis of nucleobases. The formose reaction, the most cited prebiotic source of sugars, is a complex autocatalytic set of condensation reactions which produces sugars, hydroxyl-acids and polyols from formaldehyde. It was first described by Butlerov in 1861 and

mechanistically rationalised by Breslow in 1959.[55] The reaction is initiated when formaldehyde and glycoladehyde react to produce glycerdaldehyde. A series of aldol and retro-aldol condensations proceed subsequently, in which a glycoladehyde molecule is formed, which then serves as an initiator for the reaction, making it autocatalytic. This reaction has been carried out in the presence of different mineral surfaces and inorganic species in order to alter the final product distribution and their relative stabilities.[56–58] Generally, the products obtained in this reaction are a mixture of tetroses, pentoses and hexoses, in which ribose only represents as very small fraction. Since the reaction is typically carried out under alkaline conditions, the produced sugars can decompose into tar. This is generally known as the 'asphalt problem', in which the decomposition of carbohydrates in the formose reaction does not allow sufficient accumulation of pentoses (such as ribose) to further react to form nucleosides.[59] Recently, Benner *et al.* managed to circumvent this problem by the addition of borate, which slows down sugar degradation and promotes the accumulation of a small subset of sugars (including ribose).[60]

1.2.2.2 The Nucleoside Problem

The difficult pathway to obtain the canonical nucleobases and ribose efficiently is succeeded by a greater synthetic challenge: linking them to form nucleosides.

The most rational way to synthesise a nucleoside is by reacting ribose with a nucleobase to form a glycosidic bond. In relation to the regiochemistry of the glycosidic bond, natural nucleosides react via specific nitrogen atoms in the nucleobase N(9)-position for purines and N-(1) position for pyrimidines. Given that primary amino groups in both purines and pyrimidines can also react, obtaining the right regioselectivity is rather difficult.

In 1972, Orgel achieved the formation of β-adenosine by dehydration of nucleobases and ribose at elevated temperatures in the presence of inorganic phosphates and Mg^{+2}.

Whilst these results were initially exciting, the major product corresponded to N(6)-ribosyladenine, and very low yields were obtained (up to 3%) for β-adenosine and (up to 9%) for β-guanosine.[61] Also, this reaction did not work for pyrimidine nucleobases. Due to the low regiospecifity, and the inability to produce pyrimidine nucleosides, the formation of nucleosides under abiotic conditions has been known ever since as 'The Nucleoside Problem'.

Since the RNA world theory was postulated, the limitation in both producing ribose and nucleosides efficiently under abiotic conditions has constituted the main argument against the plausibility of this theory.

However, tirelessly efforts have been put into circumventing the problems associated with the bottom-up synthesis of nucleotides. Whilst a classical approach to link the three submonomeric nucleotide constituents had only been taken, a new strategy has recently shed light on this problem. In this case, instead of trying to couple both the nucleobase and the ribose under dehydration conditions, the strategy was based on 'constructing' the nucleobase on the sugar.[61]

In 2009, Powner and Sutherland achieved the efficient synthesis of pyrimidine ribonucleotides by using this approach.[62] In this synthesis, the formation of the nucleotide proceeds via a series of very specific step-wise additions of prebiotically available starting materials, in which the formation of an amino-oxazoline intermediate facilitates the formation of the final product. These results offered an elegant solution to a problem that prebiotic chemists had been trying to solve for the last 50 years.

However, as it is common in this area of research, many objections were raised against the plausibility of this reaction. Benner pointed out that their alternative synthesis requires human addition (at the right times) of high concentrations of two carbohydrates, glycolaldehyde and glyceraldehyde.

Since Benner recently reported the stabilised formation of ribose in the presence of borate, he argued that Sutherland's starting materials would be too reactive to accumulate prebiotically.[59] Also, the reaction developed by Powner and Sutherland only produces pyrimidine nucleotides, and not their purine equivalents.

Coincidentally, when this section of the introduction was being written, an article by Becker *et al.* in *Science* reported a highly-efficient formation of purine nucleosides from abiotically available starting materials.[63] In this case, the nucleoside forms by reacting formamidopyridine (which is available from prebiotic molecules found in comets) with a sugar. This synthesis is not as restrictive as the one described by Powner and Sutherland, and high yields for purine nucleosides are obtained. Whilst these results remove one of the major objections against an RNA-based origin of life scenario, the conditions used do no match to the ones reported by Powner and Sutherland, meaning that the formation of both purine and pyrimidine nucleosides could not happen simultaneously.

1.2.2.3 Alternative nucleotides

Despite the recent breakthroughs in the synthesis of nucleotides, the classic model based on assembling their different constituent parts still faces a lot of unresolved problems. Given that most reactions aiming to produce the constituents of modern nucleotides often produce alternative nucleobases and sugars, could it be possible that the first RNA building blocks were different to the modern ones?

This possibility was first proposed by Joyce *et al.*, when it was postulated that RNA was preceded by a polymer constructed from flexible, acyclic nucleotides that were readily synthesised under abiotic conditions.[64]

Following this idea, Miller *et al.* explored the possibility to form nucleosides by reacting an alternative nucleobase (urazole) with ribose.[65] Surprisingly, the corresponding nucleoside formed in very high yields at moderate temperatures. However, the weak base pairing of urazole with adenine questions the plausibility of these heterocycles as modern nucleotide ancestors.

The challenge was then to find an alternative nucleobase that was capable of reacting efficiently with ribose to form a nucleoside[66,67], and also with enough base-pairing capacity to support template-dependent replication processes. A team of researchers led by Nick Hud has recently reported two plausible prebiotic heterocycles, melanine and barbituric acid, that form glycosidic linkages with ribose and 5-ribose phosphate in water to produce nucleosides and nucleotides in good yield.[68] Also, these nucleosides and nucleotides are capable of base-pairing in water to form non-covalent supramolecular assemblies. This work aims to understand whether a pre-RNA molecule, readily formed by abiotically available non-canonical nucleotides, could gradually evolve into modern RNA through the substitution of their building blocks. Whilst non-covalent assemblies are obtained from these alternative nucleotides, the next challenge would be to link them into a pre-RNA polymer.

1.2.2.4 Exploring new scenarios

The synthesis of adenine from ammonium cyanide in 1961 paved the way to explore plausible pathways to make nucleotides and RNA from very simple compounds. However, as the synthesis of some nucleobases and, especially nucleosides became practically

impossible, prebiotic chemists sought to find alternative ways to make those compounds, which would not necessarily agree with the pre-established ideas on prebiotic plausibility.

Recent work by Powner and Sutherland[62] has generated a controversy amongst researchers that understand the abiotic formation of RNA as a continuous process from simple precursors, without human intervention. Whilst Sutherland's work presents an elegant solution to the so called 'Nucleoside problem', the fact that it requires introducing starting materials with very high temporal precision has been criticised by those who think that such precise processes could have never happened in a prebiotic environment. Also, Sutherland has recently shown that ribonucleic acids, amino acids and lipids can all be simultaneously produced as a consequence of hydrogen cyanide and hydrogen sulphide photochemistry.[69]

This work involves a complex network of independent reactions and could not be performed in 'one-pot' conditions, but the author has proposed a geochemical model in which the three-atom precursors (HCN and H_2S) would react in different micro-environments as a result of a post-meteoritic scenario to yield the main components of the three subsystems of life (metabolism, genetics and compartments). The idea that the precursors of the three subsystems could have been synthesised simultaneously by connecting independent reactions conflicts the classical approach in which one of the subsystems originated first (RNA first or metabolism first ideas).

HCN is one of the most studied chemical precursors of biomolecules: it is involved in the formation of amino acids via the Strecker synthesis, as well as in the formation of nucleobases. However, in recent years, there has been an increasing interest in the chemistry of formamide. As opposed to HCN, formamide can act as a solvent, as well as a reactant for the synthesis of a variety of biochemical compounds. Formamide results from the hydrolysis of HCN, and recent studies suggest that it is a ubiquitous molecule in the Universe.[54] Initially, the potential role of formamide in prebiotic chemistry was considered to be limited since preliminary studies only showed the formation of a small number of heterocycles including adenine. However, heating formamide in the presence of different catalysts of terrestrial and meteoritic origin yields complex combinations of nucleobases, amino acids, sugars, amino sugars and condensing agents (Fig. 6).[70–73] More interestingly, the formation of acyclic nucleosides was achieved by simply heating formamide in the presence of TiO_2.[74] The resulting acyclonucleosides can be further phosphorylated in the presence of a phosphate source to yield 2',3'-and 3',5'-cyclic nucleotides.

Despite obtaining complicated mixtures of acyclonucleotides along with nucleobases, the simplicity of this-one pot reaction differs strikingly with the multi-step syntheses described by Sutherland[69] and Carell.[63]

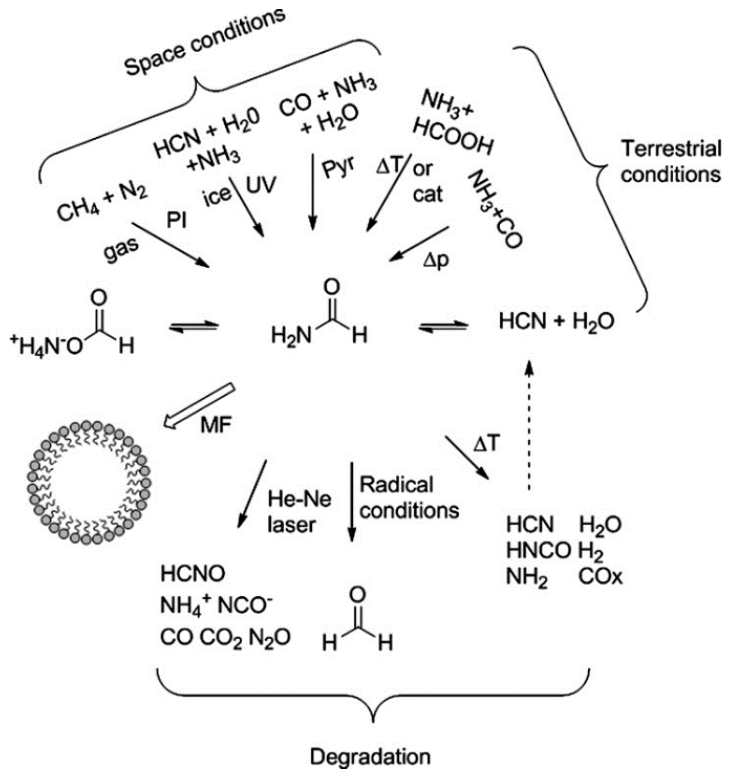


Figure 6: The basic prebiotic chemistry of formamide. PI: Photon irradiation. UV: Ultraviolet irradiation. Pyr: Pyrolysis. ΔT and ΔP: High temperature and/or high pressure. Cat: Reaction performed in the presence of catalysts. MF: Micelle formation. Adapted from Elsevier, Copyright © 2011 Elsevier B.V. All rights reserved.[71]

A recent study illustrates how an extremely rich variety of relevant prebiotic compounds can be obtained when liquid formamide is irradiated by high-energy proton beams in the presence of powdered meteorites.[75] The products obtained were amino acids, carboxylic acids, nucleobases, sugars and most notably, four nucleosides: cytidine, uridine, adenosine and thymidine. Traditionally, the laboratory synthesis of prebiotic compounds has been restricted by preconceived ideas on early earth conditions; however, new studies suggest that the exogenous synthesis of prebiotic compounds could have constituted a significant source of life's building blocks.

A large amount of evidence supports that numerous biochemical monomers can be abiotically synthesised, and that similar processes could have occurred either under early earth or space conditions. However, how might abiotically available monomers react to produce the two types of polymers that are essential to life?

1.3 Synthesis of biopolymers

Most of life's biopolymers are synthesised when a water molecule is removed to form ester bonds in the case of nucleic acids, glycosidic bonds in the case of saccharides and peptide bonds in proteins. In modern cells, these condensation reactions happen in water, despite being thermodynamically unfavourable in aqueous media. This happens due to complex and highly evolved processes which involve the presence of enzymes and energetically driven reactions by the hydrolysis of pyrophosphate bonds.

These highly evolved processes could not have taken place in the early stages of chemical evolution. Given the set of available abiotic monomers, the next step would be the synthesis of their respective polymers: peptides and nucleic acids. However, a plausible abiotic process capable of coupling monomers into long biopolymers still remains elusive.

The next section summarises the most relevant work on the bottom-up synthesis of both peptide and nucleic acid polymers.

1.3.1 Peptide synthesis

Peptide bond formation is thermodynamically unfavorable in aqueous solution: the energy required to form a peptide bond in water is estimated to be between 2.5 and 3.6 kcal/mole.[76] Many possibilities have been studied in order to make peptide bond formation a downhill reaction. The most straightforward way is to perform the reaction in the absence of water. In application of Le Châtelier's principle, if the condensation between two amino acids releases a water molecule to form a peptide bond, removal of the water drives the equilibrium towards the product side.

Another mechanism to reduce the thermodynamic barrier is to increase the solution temperature. Thermodynamic calculations have shown that the polymerisation $\Delta_r G°$ of Gly into GlyGly becomes negative at more than 400 °C.[77] However, monomer decomposition processes would prevent amino acid polymerisation.

Alternatively, peptide bonds can be synthesised in the same way the cell machinery does: by coupling the thermodynamically unfavourable peptide bond formation with a thermodynamically favourable one.

This is usually accomplished with the use of coupling agents or activators. The most relevant pathways described in the literature to obtain peptides by the aforementioned strategies are discussed in the next section.

1.3.1.1 Peptide thermal synthesis

The first attempt to polymerise amino acids under abiotic conditions should be attributed to Meggy in 1953. Whilst it is not clear whether Meggy's interests were to understand how proteins could have been formed under abiotic conditions, he was the first to describe the synthesis of polyglycine by simply dehydrating a solution of glycine monomer. In his first attempts[78], Meggy obtained significant amounts of an insoluble polymer when a concentrated solution of glycine 2,5-diketopiperazine was heated at 180 °C, but no polymer was obtained when glycine monomer was used as a starting material. Following on this work, Meggy then described the formation of polyglycine from glycine monomer by adding HCl.[79] The conditions by which these experiments proceeded efficiently to produce polymers are often contradictory, and it's not clear whether all these reactions were performed in the same way.

However, Meggy noted that the rapid evaporation of a glycine solution led to no polymer formation. On the contrary, if the same solution was first refluxed for two hours, and then evaporated, the obtained yield for polyglycine formation was significantly increased. In this work, there was little analysis of the product mixtures, and the length of the polymers obtained was estimated to be slightly bigger than 10.

The formation of a glycine-alanine co-oligomer was also achieved with no other product characterisation than a positive biuret test. This is the first example where the thermodynamically uphill formation of peptide bonds was overcome by removing the water from the reaction at high temperatures. Interestingly, this article has been cited very few times despite being the first to describe the formation of peptide bonds from unactivated monomers. The reason for that might as well be that just a few years after this publication, the formation of larger 'proteinoid' species by Sidney Fox took all the attention from the scientific community.

In 1958, Fox and his colleague Kaoru Harada reported the formation of long polymers resembling proteins, which then became known under the name of 'proteinoids'.

To make these polymers, a dry mixture of amino acids was heated up to 180 °C under completely anhydrous conditions. In his first publication[80], and inspired by the inability from Meggy to thermally co-polymerise different amino acids, Fox reported the formation of glycine-glutamic acid co-oligomers, which he initially characterised as linear peptides.

In these experiments, a higher glycine to glutamic ratio was used, and the formation of polymers with an average molecular weight of 15000 Da was described. Shortly after, he published an article in *Science*[35] where the formation of proteinoids including 13 different amino acids was achieved. In this case, however, 80% of the initial amino acids composition included aspartic acid and glutamic acid. According to Fox's observations, pyroglutamic acid (which readily formed upon high temperatures) was liquid in the presence of other amino acids at 170 °C and therefore could act as a solvent.

A few years later, Fox and his collaborators discovered that proteinoids were able to self-assemble into microscopic spherical balls, which they referred to as proteinoid microspheres (see Fig. 7).[81] These spherules could still maintain their integrity after a few weeks, as well as interact with other spherules by coalescence or aggregation. Fox also proposed that these spherules could serve as primitive compartments.

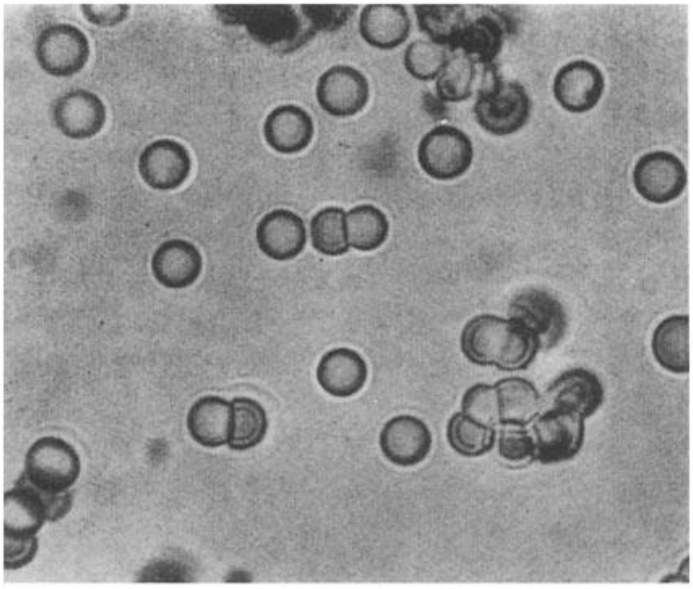


Figure 7: Uniform proteinoid microspheres of bacterial size range.Adapted from Springer, Copyright © 1995, Kluwer Academic Publishers[82]

For a long time, protenoids were considered to be the predecessors of modern proteins. Fox reported in many occasions how these species had various catalytic activities, including the ability to catalyse the formation of peptide bonds in the presence of ATP.[83]

Other researchers then started to expand Fox's work interested by the wide range of catalytic and structural properties that resulted by changing the relative amino acid proportions in the thermal polymers.[84–86] A thorough analysis of the structure and catalytic activities of thermal polyaminoacids was described by Dose in 1974[87].

With the discovery of the catalytic function of RNA and the birth of an RNA-world based scenario, proteinoids lost the relevance they initially had.

However, this did not stop Fox to advocate for a protein-first scenario, where proteinoid microspheres would be the first-living cells as well as proto-neurons. These statements were usually based on observations that lacked scientific rigor, such as proposing proteinoid microspheres as the first cells due to their resemblance to microfossils, amongst other examples.

Fox's statements in the latest stages of his career contributed to the underestimation of his scientific achievements. The thermal synthesis of proteinoids is nowadays completely dismissed as a way of forming peptides under abiotic conditions.

Several reports in the literature describe Fox's proteinoids as a mixture of products containing mostly ester bonds with a very small fraction of peptide bonds.[40] Whilst there is no characterisation of such products with the use of modern analytical techniques, Dose[87] performed an exhaustive analysis of proteinoid's composition back in 1974.

Another independent research group studied in detail the nature of these products by NMR in 1976.[88] These analyses showed that proteinoids contain 40% to 60% less peptide bonds than proteins and that side reactions during the condensation lead to largely unknown products which are linked to the amino acids other than peptide bonds. Also, given the presence of a high proportion of aspartic acid, glutamic acid or lysine, β, γ or ε peptide bonds are more easily formed than α-peptide bonds. The formation of peptide bonds by melting dry amino acids at high temperatures was then abandoned in favour to chemical systems that worked in water at moderate temperatures.

However, some of Fox's later experiments reported that proteinoid formation was achievable at lower temperatures (100 °C) in the presence of phosphoric acid[89]. It was first suggested that phosphoric could act as a solvent, as a dehydrating agent and also as an acid catalyst. The mechanism by which the presence of phosphoric acid could catalyse the formation of proteinoids was not described. Shortly after, he published new results where polyphosphoric acid was used instead. In this case, it was suggested that the formation of

phosphoryl-amino acid intermediates could be the reason why this reaction proceeded efficiently at much lower temperatures, but no evidence was shown to prove this hypothesis due to their analytical limitations (note that the first NMR units were still being developed when this work was published).

Nevertheless, whilst studying the effect of phosphoric acid in proteinoid formation , he performed the same reaction in the presence of sulfuric acid and did not observe the formation of polypeptides, despite the similar dehydrating properties between sulphuric and phosphoric acid. In most proteinoid experiments, more than 50% of the starting material corresponded to amino acids electrically charged side chains. Recent work by Varfolomeev[90] also reported the thermal copolymerisation between amino acids with electrically charged side chains (Asp, Glu, Lys and Arg) in the presence of sulfuric acid. The main products obtained in these reactions are peptide derivatives containing succinimide or pyroglutamic acid residues in high proportion.

It is well-stablished that amino acids with a carboxylic acid (aspartic/glutamic acid) side chain readily homopolymerise under thermal conditions in the presence of an acidic catalyst to yield polymers mainly composed of cyclic imide bonds[91]. In the synthesis of polyaspartic acid, aspartic acid is heated at high temperature which results in the formation of polysuccinimide. Then, by reacting the polymer with NaOH, the cyclic succinimide rings are cleaved, which result in the formation of a linear polyaspartate polymer. However, the resulting polymer contains 70% of the amide bonds formed from β-carboxyl groups and 30% from α-carboxyl groups (Fig. 8).

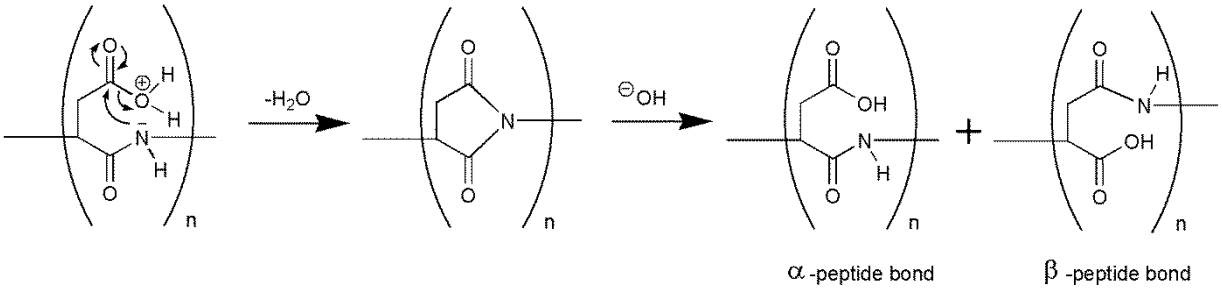


Figure 8: Schematic representation of polyaspartic thermal synthesis, resulting in the formation of α and β peptide bonds.

1.3.1.2 Peptide formation and inorganic catalysis

During the 70s, there was a growing interest in the use of minerals to catalyse the formation of peptides. In particular, clay minerals were not only attractive because of their ability to catalyse a wide range of organic reactions, but also because of their probable contribution to prebiotic organic complexification due to a limitless array of catalytic sites. This idea was first proposed by Bernal in the 1950s and later named as 'Polymerization on the rocks' by Orgel.[92]

The catalytic effect of mineral surfaces has been recently reviewed in detail by Lambert.[93] Many studies have focused on the interactions of amino acids with minerals with the aim to demonstrate that the adsorption of amino acids on a mineral surface could favour polymerisation.

Nonetheless, a vast number of different parameters determine the adsorption rate of amino acids on mineral surfaces (molecule size, charge, solubility, mineral charge), as well as experimental conditions (pH, temperature, presence of inorganic ions). Lambert concludes that even if amino acid polymerisation is favoured in the absorbed state, the kinetics might be quite slow. He also states that if amino acid adsorption is coupled to a process that decreases the activity of water in the system, polymerisation reactions can be made favourable. Whilst a rational explanation of this process has been described recently by Lambert, the first example dates back to the 1970s, when Lahav *et al.* obtained measurable amounts of glycine oligomers by repeating wetting-drying cycles in the presence of kaolinite and bentonite at 80 °C.

A few years later, a new approach to reduce the activity of water to promote peptide bond formation was proposed by Bernd Rode.

He observed that highly concentrated solutions of NaCl could effectively act as a condensation reagent due to the strong binding energy of Na^+ ions to the hydration shell. High concentrations of sodium ions act as a dehydrating agent to promote the condensation reaction by reducing the thermodynamic barriers for peptide bond formation. At the same time, Rode calculated theoretically how metal complexes could help to lower the kinetic barrier of condensation reactions between organic ligands. By merging these two findings, Rode was able to form peptide bonds in aqueous solution by simply adding a high concentration of NaCl as a dehydrating agent and Cu^{+2} to form complexes with the amino acids.

The complex formation catalysed the condensation between the unprotonated amine from one amino acid and the carboxylic acid of the adjacent amino acid. The chloride ion from the highly concentrated NaCl salt also acts as a stabilising ligand in the reaction mechanism (see Fig. 9). Different divalent ions were tested but only Cu^{+2} was found to efficiently promote the formation of peptide bonds. The reaction also proceeded when Mo^{+6} was used instead, but only diglycine could be obtained. Interestingly, the reaction prefers α-amino acids over their β and γ analogues, which can be explained by the more favourable 5-membered ring that forms when an α-amino acid chelates a Cu^{+2} ion, as opposed to the less stable 6 and 7 membered rings that form when β or γ- amino acids are used, respectively.

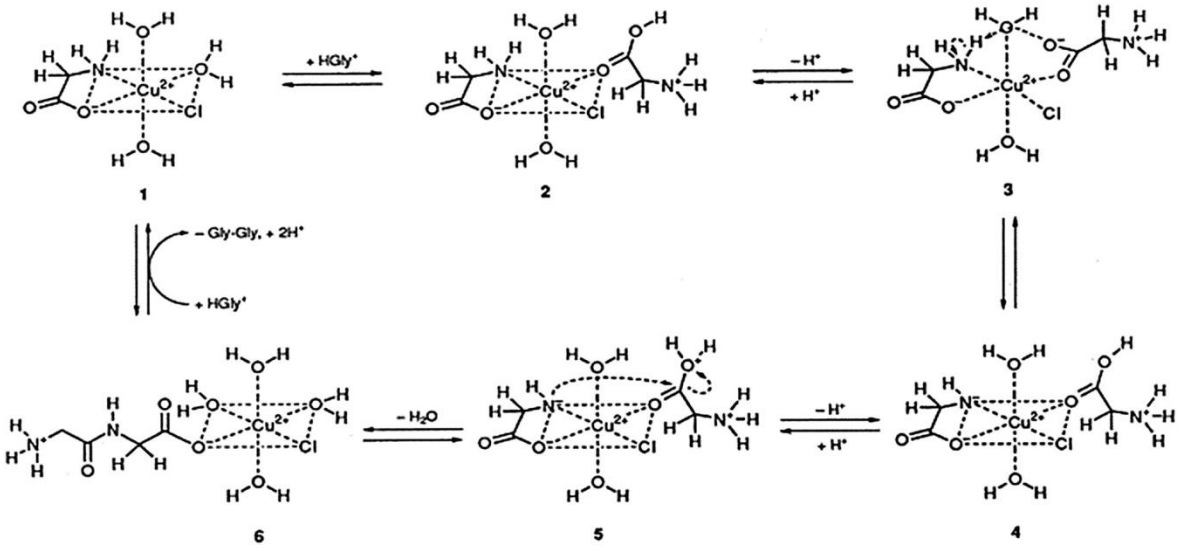


Figure 9: Reaction mechanism for peptide formation of glycine monomers via the salt induced peptides formation reaction. Adapted from the Royal Society of Chemistry, Copyright © 2012, Royal Society of Chemistry.[40]

Also, in the presence of Montmorillonite and other clays, yields were better and larger oligomers were formed.

The role of the clay mineral in this reaction could have been protecting the oligomers against hydrolysis, as well as connecting small peptides on its surface to form larger species. It was suggested that in these experiments, the presence of specific amino acids could catalyse the formation of homo-dipeptides of other amino acids.

This finding could explain why in the SIPF (Salt-Induced Peptide Formation) reaction there are some clear preferences for certain amino acid linkages, as opposed to a statistical mixture of all possible combinations.

Excited by these results, Rode attempted to correlate the most frequent binary amino acid linkages in the SIPF reaction with those present in membrane proteins in archaea and prions, and concluded that there might be a 'fingerprint' of the SIPF reaction in the composition of the first proteins.

The SIPF is described by the author[40] and also in the literature[94] as 'the most simple and universal explanation for peptide formation under realistic prebiotic conditions'. However, only short glycine oligomers (up to 6-mers) in very low yields can be obtained when successive hydration-dehydration cycles are performed.[95]

In accordance to the experiments shown above, peptide formation has also been observed using Langmuir-trough methods at the air-water interface with Cu^{+2} molecules as catalysts. The air-water interface has been proposed to be important in prebiotic chemistry because it provides a unique environment through chemistry through its ability to concentrate and align biochemical precursors[96,97]. Therefore, aerosols have been suggested as prebiotic chemical reactors for the condensation of amino acids.

1.3.1.3 Peptide hydrothermal synthesis

It has also been proposed that oligopeptides could have been formed in submarine hydrothermal systems (SHSs). Given the relevance that SHSs acquired as a potential scenario where life could have emerged[31], an extensive amount of research has focused on the formation of peptides under hydrothermal conditions. Hydrothermal experiments involve the heating (usually >200 °C) of aqueous solutions of amino acids in a confined hydrothermal reactor at high pressures.[98–100]

Several experiments have reported the formation of short peptide oligomers by simulating the conditions of SHSs in the presence of different catalysts such as Cu^{+2} ions[101], fatty acids[102] or clays.[103]

An extensive review by Cleaves *et al.* discusses the plausibility of hydrothermal vent systems to promote the formation of peptide bonds.[104] Cleaves notes that all the experiments that simulate SHSs start with very high concentrations of amino acids, which

would be unlikely to occur in a plausible geochemical scenario. Also, at such high temperatures, peptide bond formation would compete with monomer degradation, which would shift the monomer-polymer equilibrium towards net depolymerisation.

After a thorough investigation of a series of conditions (pH, temperature, concentration), Cleaves concludes that alternative cooler environments such as evaporative basins are more likely for the concentration and polymerisation of organic molecules under primitive conditions. Nonetheless, a few examples on the formation of peptides under SHSs conditions have been recently reported, despite being questioned previously.[103,105] As expected, none of these experiments have produced oligomers longer than 6-mers, and the reported yields did not exceed 1%.

The examples mentioned above represent the most relevant work on the synthesis of peptide bonds from unactivated amino acids either via the removal of water, by using high concentrated salt solutions as dehydrating agents or under hydrothermal conditions to promote the condensation reactions.

1.3.1.4 Peptide synthesis using activated amino acids

Parallel to the aforementioned examples, significant work has focused on the synthesis of peptides by using coupling reagents.

The first example was shown by Ponnamperuna[106], who suggested that cyanamide (which can be abiotically synthesised from ammonium cyanide) could be used as a condensing agent to promote the formation of peptides from unactivated amino acids. His preliminary results showed the formation of peptide dimers and trimers when amino acid solutions were dehydrated at 80 °C in the presence of cyanamide, ATP and $MgCl_2$ and exposed to UV irradiation.

Follow-up work was performed by Oró, who achieved the formation of peptide oligomers up to tetramers under similar conditions.[107,108] Calvin observed a similar effect when dicyandiamide was used instead under dehydrating conditions.[109]

However, the use of cyanamide and some of its derivatives did not produce peptides with sufficient length to fold into well-defined strcutures with endowed catalytic activities. Recently, Danger *et al.* have revisited the use of cyanamide as a prebiotic condensing agent

and discovered that it can promote peptide elongation through the formation of the cyclic intermediates 5(4H)oxazolones.[110]

A different approach was taken in the 1990s when Orgel decided to model a prebiotic polymerization reaction by using condensing reagents such as carbonyl diimidazole (CDI) and 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDAC). In the presence of these condensing reagents, Orgel and co-workers managed to obtain a high degree of polymerisation (up to 55-mers) upon adsorption of mineral surfaces (hydroxylapatite and illite).[111] Whilst these results show the longest peptide oligomers that have been synthesised in a prebiotic model reaction, CDI and EDC could be hardly considered as prebiotic reagents. In the case of water soluble carbodiimides, such as EDC, high polymerisation degrees can only be obtained with aspartic acid. Nonetheless, it was observed that in the presence of these condensing reagents, the polymerisation reaction proceeded via the formation of an N-Carboxy anhydride (NCA) intermediate.

NCAs were first reported by Hermann Leuchs in 1906[112], and since then, these cyclic and reactive amino acid derivatives have been used for the formation of polypeptides by ring-opening polymerisation. This type of activated amino acids has the advantage that their ring structure combines the activation of the carbonyl group while the amino group is protected.

The facile synthesis of homo- and copoly-amino acids by ring opening polymerisation of NCAs served as models of natural polypeptides and proteins, which helped to understand the relationships between primary and secondary structures. For the last 50 years, a significant amount of work has been put into developing new strategies to efficiently control the polymerisation reaction and the final product structure, with special emphasis on the synthesis of polypeptides with various architectures (diblock, triblock, multiblock sequences, star-shaped or dendritic structures) and their possible applications as drugs or drug carriers. A thorough review on the development of NCA chemistry during the last century was recently written by Kricheldorf.[113]

However, the role of NCAs as activated monomers for the prebiotic synthesis of polypeptides has only been considered during the last 30 years.

As mentioned before, NCAs were found to be reaction intermediates in the formation of polypeptides when unactivated amino acids were reacted with CDI.[111]

Since CDI has never been considered as a prebiotically plausible reagent, alternative ways of producing NCAs under prebiotically accepted conditions have been explored (Fig. 10). The formation of oligopeptides through NCA intermediates was observed when amino acid derivatives (amino acid esters[114,115] or amino acyl adenylates[116,117]) were treated with CO_2 or bicarbonate solutions.

An alternative approach was explored by Commeyras *et al.*, who found that carbamoyl amino acids (CAAs) easily react with NO/O_2 gas mixtures to form NCAs in quantitative yields[38,44,118–121]. This reaction requires a concentration process by evaporation to ensure an efficient exposure of CAAs to NO/O_2 gas mixtures from the atmosphere, as well as to prevent the hydrolysis of the formed NCAs in water.

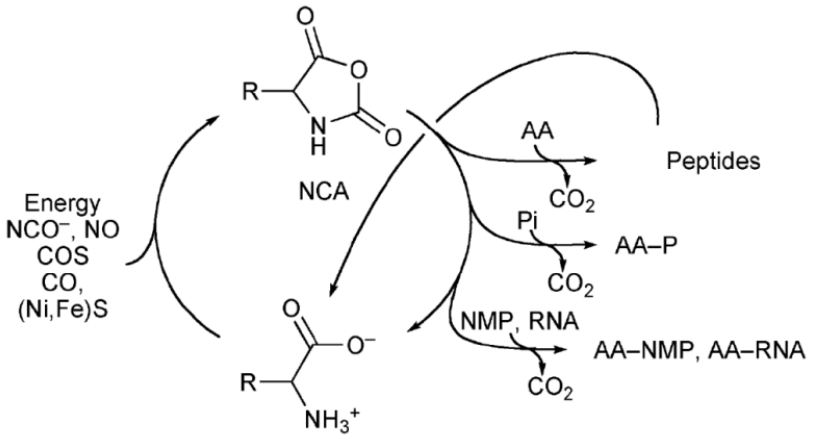


Figure 10: Pathways for the formation of NCAs and further reactions including polymerization and interactions with inorganic phosphate (Pi), nucleotides (NMP), and RNA. Adapted from the Royal Society of Chemistry, Copyright © 2012, Royal Society of Chemistry.[122]

Based on the requirements of this reaction, Commeyras proposed a scenario where successive hydration-dehydration cycles of CAAs in the presence of NOx gases could produce heterogeneous peptide species via the formation of NCAs intermediates.

In the so called 'Primary Pump scenario', it is expected that successive hydration-dehydration cycles will promote the accumulation, elongation and complexification of peptides when the hydration cycle is short enough to prevent the hydrolysis of NCAs (Fig. 11). Under a continuous supply of amino acids and energy, this model aims to represent a protometabolism in which homochiral peptides could emerge from a racemic amino acid mixture. Preliminary experiments on this model have not succeeded to produce peptides longer than 5-mers. However, sequence selectivity was observed when only 6 out of 32 possible sequences were obtained for the produced 5-mers.[118]

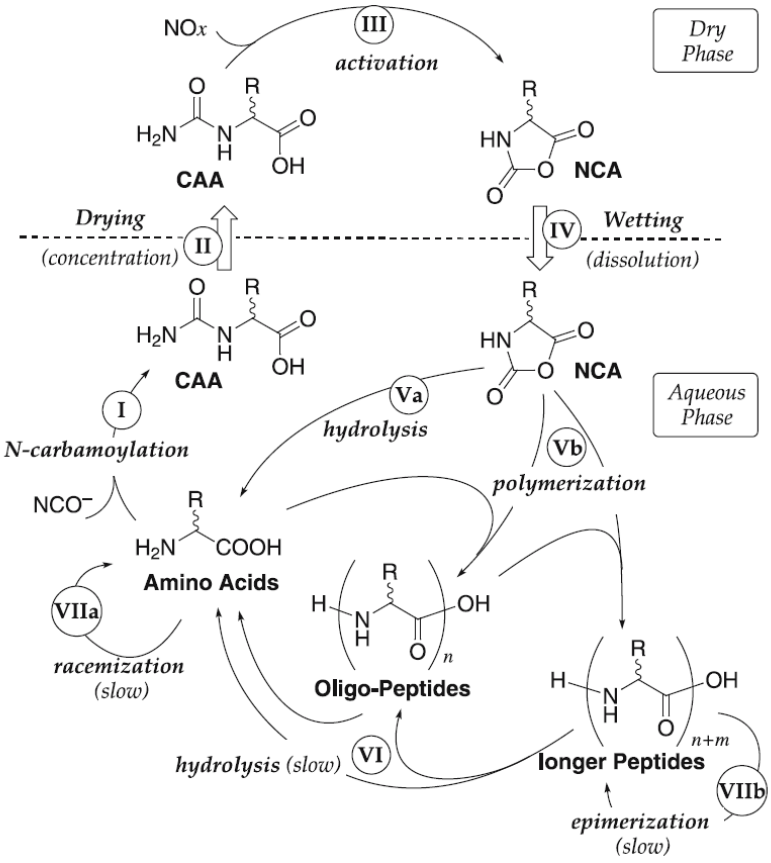


Figure 11: The Primary Pump, a peptide-based protometabolism scenario involving the following steps: I: amino acid N-carbamoylation; II: concentration through drying; III: NOx-mediated CAA activation; IV: dissolution through watering (by e.g. sea water); V: NCA reaction in aqueous phase (Va: NCA hydrolysis; Vb: condensation with AA or peptide); VI: slow hydrolysis of peptide bonds; VII: α-carbon epimerization (VIIa: of amino acid and CAA; VIIb: of peptide residues). Additional steps corresponding to peptide N-carbamoylation/nitrosation have not been mentioned for the sake of clarity. It is worth mentioning that although the N-carbamoylation of peptides renders them unreactive towards NCA, this is reverted by NOx-mediated nitrosation, thus keeping peptides within the polymerization process. Reprinted with permission from Springer, Copyright © 2005, Springer.[38]

The most recent work on the formation of peptides from activated amino acids proceeds through a different pathway, in which the amino acid is activated by carbonyl sulphide (COS), which is present in volcanic gases in very low proportions. In this reaction, thiocarbamates are formed, which are subsequently converted into NCAs in good yields.[123]

In summary, the main problems associated to the formation of peptides from activated amino acids is that the only reaction pathways capable of producing very long peptides (i.e. 55-mers in the presence of mineral surfaces) are not attractive since the prebiotic significance of the condensing agents used is doubtful. On the other hand, alternative strategies to produce activated forms of amino acids such as NCAs or 5(4H)-Oxazolones with presumably more prebiotically plausible reagents only produce short peptides.

The main challenge on the abiotic synthesis of peptides is to find a simple reaction pathway that produces polypeptides long enough to fold into catalytically active structures. Whilst many efforts are being put into solving this problem, the latest examples on abiotic synthesis of peptides still focus on the effect of mineral surfaces[124] or hydrothermal conditions[104], which usually do not produce peptides longer than 3-mers. However, a few novel approaches are exploring alternative ways of overcoming this problem. Very recently, polyglycine oligomers up to 16 units long were obtained by successive feeding of monomers from the vapour phase on a SiO_2 surface.[125] This study is the first to report the formation of polypeptides on a surface by successive monomer feeding and the formation of species that presumably possess both helical and β-sheet structures. Whilst this result is encouraging, and long species are formed, the formation of glycine homo-oligopeptides is only described. Alternatively, a model reaction in which successive wet-dry cycles drive the formation of oligoesters has been suggested as a possible mechanism for the formation of the first biopolymers.[126]

1.3.2 Nucleic acid synthesis

The synthesis of RNA oligomers from unactivated nucleotides has many resemblances to the synthesis of peptides. Since the formation of a phosphodiester bond requires the loss of a water molecule, and is also an uphill reaction with a free energy of 5.3 kcal/mol, very similar strategies to those performed to synthesise peptide bonds have been explored.

When starting from unactivated nucleotides, the most rational approach of heating solutions of monomers until complete evaporation has failed to produce long oligoribonucleotides.[127] Hydrothermal syntheses have also been explored, but only short oligomers have been obtained due to the high hydrolysis rates at elevated temperatures[128,129].

Since unactivated routes to form RNA oligomers were unsuccessful, the use of condensing agents and activated forms of nucleotides as starting materials were also explored. Condensing reagents derived from abiotically available molecules (cyanamide, water soluble carbodiimides, etc.) only produced dimers and trimers.[130,131] Likewise, the use of activated phosphates (whose prebiotic relevance is under debate) was able to promote the synthesis of short oligoribonucleotides.[132]

On the other hand, most attempts to synthesise oligoribonucleotides involve the use of activated forms of nucleotides such as 5-phosphorimidazolides. The condensation of these activated monomers in aqueous solution yields a mixture of 5',5'-pyrophosphate-, 2',5'-phosphodiester-, and 3',5'-phosphodiester-linked adducts, in order of decreasing abundance.[133] The poor regioselectivity of this reaction led researchers to explore other conditions in which the formation of the 3',5'-phosphodiester bond was favoured (i.e. using metal catalysts or in eutectic solutions).[134,135]

However, one of the most relevant synthesis of RNA oligomers was described by Ferris *et al.* in numerous publications[111,136–138], in which oligomers up to 40-mers were obtained when minerals such as Montmorillonite or hydroxylapatite were reacted to 5-phosphorimidazolides activated nucleotides. Most remarkably, the majority of the phosphodiester bonds obtained were 3'-5', showing that interactions with the mineral surfaces were affecting the regioselectivity of the reaction. This work contributed to consolidate the role of mineral surfaces in prebiotic chemistry. As in the case of peptides, in which highly reactive NCA intermediates are difficult to be produced by prebiotically available precursors, the same problem is encountered for activated nucleotides.

Strategies to synthesise activated nucleotides involve the addition of nitrogen-containing heterocycles (such as imidazole and pyrimidine derivatives), phosphorimidazolides being the most common activated form. It is not clear whether these species could have existed in sufficient amounts on the early Earth to promote the formation of long RNA polymers.

Recently, a different approach that does not require the use of additional activating species has shown the formation of very long RNA oligomers. In previous work, a team of researchers led by Ernesto Di Mauro managed to overcome the so called 'nucleoside problem' by being able to synthesise different nucleosides by exposing formamide to a high-energy proton beam or UV radiation.[74,75] Some of the products of this reaction include nucleosides and acyclonucleosides.

The reaction of these products with different phosphate minerals yield a variety of nucleotides which include phosphorylation at any of the possible positions of the sugar moiety: 2', 3', 5' and 2',3' or 3',5' cyclic.[139] In a model reaction, it was then found that cyclic nucleotides (3',5'-cAMP and 3',5'cGMP) readily polymerise in aqueous solution by base catalysed trans-phosphorylation of stacked monomers to produce up to 120-mers[140].

The reaction proceeds at different stages: first, oligomerisation of 3′,5′-cGMP produces up to 25-nucleotide-long RNA molecules, whilst only up to 8-mers are obtained for 3',5'-cAMP; secondly, RNA extension occurs by polymerisation on pre-existing oligomers and finally, ligation between pre-formed RNA oligomers produce molecules >100 molecules long. Nonetheless, very low yields are only obtained, and several concerns about the reproducibility of this work have been raised[141].

In 2014, Morasch *et al.* decided to experimentally test the validity of this reaction by using a different approach.[142] Instead of performing the reaction in aqueous solution at elevated temperatures (40 °C to 80 °C), they managed to obtain poly-G RNA strands up to 40-mers by simply drying a solution of 3'-5' cGMP. In this case, it was shown that even in the presence of activated nucleotides, removal of water significantly improved the yields and the reproducibility of the reaction.

Similar experiments which aim to synthesise long RNA oligomers by water removal have also been recently described by a team of researchers led by Dave Deamer. Instead of using cyclic forms of nucleotides, Deamer *et al.* have shown the formation of long RNA-like oligomers when unactivated nucleotides are exposed to successive hydration-dehydration cycles in the presence of lipid molecules.[143,144]

During the dehydration step, the reactants are concentrated and organised between lipid lamellae in such a way that ester bond synthesis is promoted. With this initial approach, very low yields are obtained <1%.

A follow-up study which involves the use of monovalent salts achieved up to 55% yields when feeding experiments were performed, in contrast to the much lower yields obtained when lipids were used.[145] In this particular experiment, mononucleotides are concentrated and organised within the eutectic phases that form upon crystallisation of the monovalent salts. The oligomers formed are referred as RNA-like oligomers since it is not known the relative ratios between 2',5' and 3',5' phosphodiester linkages. Several analytical techniques (gel electrophoresis, nanopore analysis, hyperchromicity tests) reveal that the products are linear polymers resembling RNA, and that secondary structures with duplex character can form when equimolar mixtures of AMP and UMP are employed.

1.3.3 The importance of homochirality

In most prebiotic experiments, the synthesis of life's building blocks results in the formation of both left-handed and right-handed forms. However, in all living beings, both peptides and nucleic acid are made of enantiomerically pure components: L-amino acids and D-sugars, respectively. This is very important since proteins comprising D and L amino acids cannot form well-defined tertiary and quaternary structures, whilst RNA cannot adopt functional structures if it is comprised of D and L sugars. Hence, if biopolymers were comprised of D and L monomers, they would not be able to perform their function.[146]

The main question that still remains unanswered is how the first homochiral biopolymers were formed. There are two main hypotheses: the first one suggests that long homochiral biopolymers formed after the appearance of the first living systems, whilst the second suggests that homochiral polymers were formed under abiotic conditions.[147] If we assume that the formation of homochiral biopolymers proceeded through abiotic processes, there might have been a symmetry-breaking process leading to the enantioenrichment of one type of monomer, or alternatively, it could have been possible that the enantioenrichment occurred alongside the synthesis of biopolymers. Previous work by Lahav[148] has demonstrated that by taking advantage of the supramolecular architectures formed during polymerisation reactions (β-sheets or α-helices), these can be used as templates for the generation of isotactic peptides. In this case, racemic activated α-amino acids were organised in 2D and 3D crystalline surfaces, which produced short isotactic peptides that self-assembled into β-sheet colloidal particles. Then, the β-sheet structures operate as stereoselective templates in the formation of longer racemic homochiral (isotactic) peptides.

1.3.4 RNA-peptide systems

Most research on the abiotic synthesis of both peptide and nucleic acid oligomers has only been attempted under very specific conditions, in which pure monomers were the starting material of the reaction. Very few examples can be found in the literature in which both peptide and nucleic acid systems are reacted together. These two biopolymers are so intimately linked in modern's life machinery that is difficult to conceive a scenario where they did not co-exist. Given the limitations on the oligomerisation of both peptide and nucleic acid monomers, it has been suggested that early interactions between these two would have led to emergence of early translation machinery, which then evolved into a more complex proto-ribosome.

During the translation mechanism, the enzyme catalysed formation of aminoacyl adenylates is crucial for the subsequent formation of aminoacyl-tRNAs, which then deliver the amino acid to the ribosome to grow a polypeptide chain. Assuming that no sufficiently evolved enzymes were present before the emergence of the translation machinery, how could small nucleotides and amino acids interact to generate reactive intermediates which promoted a simpler mechanism for peptide synthesis?

Initial attempts to investigate synergistic interactions between amino acids and nucleotides showed that small peptides can be formed in the presence of amino acids, ATP and cyanamide.[108]

Recent studies have focused on the development of a mechanism that enables the formation of reactive aminoacyl adenylate anhydrides, which could then further react to form RNA oligomers and peptides.

Pascal *et al.* have developed a theoretical model in which hydrolysis of aminoacyl adenylate anhydrides coupled to a metabolic pyrophosphate cycle would drive the polymerization of ribonucleoside triphosphates.[38] Whilst this theoretical approach has not been demonstrated experimentally, it has been recently found that histidine-containing dipeptides are able to catalyse the formation of phosphodiester bonds.[149]

Whilst the aforementioned studies do not specifically address the origin of coded peptides, extensive research carried out by Yarus and co-workers have shown that simple chemically-driven coded polypeptide synthesis could result from interactions between RNA oligomers and amino acids or short peptides.

It has been suggested that the association between amino acids and particular codon triplet sequences are remnants of direct interactions between amino acids and RNA before the appearance of the ribosomal machinery. This hypothesis has been further corroborated by the analysis of RNA aptamers that specifically bind to a variety of amino acids, in which at least in 8 cases, there was a relationship between the binding site of the RNA and the anticodon triplet that corresponds to that particular amino acid.[150] Also, it was later discovered that a specific RNA aptamer was capable of binding a dipeptide with higher affinity that their individual amino acids[151]. More interestingly, the aptamer binding site corresponded to the anticodon sequence of one of the amino acids. Whilst there is not still evidence about RNA aptamers that catalyse the formation of peptide bonds, further work is required to better understand the potential for RNA catalysis in the origin of translation.

1.4 Missing link between polymers and self-replicating systems

Researchers have still not succeeded to synthesise nucleic acids or proteins from simple precursors in a continuous process without human intervention. Nonetheless, recent promising discoveries (such as the formation of nucleobases and nucleosides from formamide chemistry[73], and the formation of RNA oligomers from cyclic nucleotides[152] or under wet-dry cycles[145]) have shed light on the possibility of producing genetic polymers from the bottom-up in a simple, straightforward experiment that does not require a high level of human intervention.

On the other hand, researchers have failed to produce peptides long enough to possess a well-defined structure or enzymatic activity. However, some studies suggest that even short dipeptides could have played an important role in catalysing the formation of peptide bonds and peptide nucleic acids, such as Ser-His.[153] Historically, the synthesis of nucleic acid polymers has attracted more attention since the RNA world model became widely accepted. The impossibility to make RNA oligomers from the bottom-up led to researchers to explore the problem from a top-down approach.

1.4.1 Top-down approaches

Top-down approaches aim to create the simplest version of a living cell by reducing the chemical inventory present in complex life forms. In this sense, different levels of simplifying the components in minimal artificial cells have been explored.

A team of researchers led by Craig Venter pioneers the synthesis of artificial life forms by chemically synthesising a bacterial genome and replacing it by the original one in a bacterial cell. Recently, his team has synthesised a mutant of *Mycoplasma mycoides* (one of the simplest known bacterium) with only 473 genes by removing non-essential genes.[154] Whilst this constitutes the smallest genome known in any living organism, 28% of these protein coding genes were of unknown function, and yet they were necessary to ensure the viability of the cell. This shows that we are still far from fully understanding the huge complexity that still operates within a minimal artificial cell. Alternative approaches explore the formation of a minimal cell with a much more reduced set of building blocks, in which the greatest challenge consists on coupling the self-replication of a genetic polymer to the self-division of the compartment in which it resides.

Since most of these experiments aim to produce a minimal artificial cell that agrees with an RNA-first scenario, the focus has been placed on the synthesis of an RNA replicase (an RNA molecule that can act both as a template for the storage and transmission of genetic information, and as an RNA polymerase that can replicate its own sequence) within an enclosed boundary capable of self-division.

However, whilst several advances have been made in driving RNA-catalysed RNA copying, the poor copying accuracy and the difficulty to separate RNA strands still constitute the main challenges before self-replication can be achieved.

A thorough review on the problems associated to non-enzymatic RNA replication has been recently written by Szostak.[155] Even if the formation of an RNA molecule capable of undergoing self-replicating cycles was possible, in order to constitute a living entity, it would have to carry out metabolic functions that conferred an advantage to its host compartment. In this line of thought, it has been proposed that an RNA replicase should be able to catalyse its own formation, as well as the formation of a second ribozyme that catalyses the synthesis of membrane lipids, which would then increase the growth and the subsequent vesicle division.[156]

Whilst this latter scenario still looks far from being achieved, incremental progress is being made in integrating the non-enzymatic templated directed synthesis of a genetic polymer within fatty acid vesicles. Nonetheless, there are still critical problems that remain to be solved in the construction of RNA-based protocell analogues. Szostak, who has dedicated most of recent years developing RNA-based protocell models, clarifies that given the difficulties encountered in building such systems, it could have been possible that primitive versions of RNA would be easier to synthesise and replicate than their modern equivalents.[155]

The inability to construct simple living protocells by reducing the complexity of their components leads to focus our efforts in the bottom-up synthesis of biopolymers capable of replication and catalysis, without constraints based on RNA-only or peptide-only systems.

1.4.2 Prebiotic Systems Chemistry

The classical approach in prebiotic chemistry has focused on very specific syntheses of precursors and their respective polymerisation under conditions constrained by hypotheses on geochemically plausible scenarios. Recently, the field of Systems Chemistry has emerged to give response to the problems encountered with the classical approaches. In this sense, Systems Chemistry aims to provide new fundamental insights by studying multiple variables simultaneously, and the properties that emerge from the interaction between different chemical components within a complex mixture. The recent advances in analytical methods (HPLC-MS, multidimensional NMR) have allowed the quantification of individual components within complex mixtures. This could help to elucidate the evolution of mixtures over time, and the role played by their individual components.

Whilst Systems Chemistry is very broad and can be applied to many chemical systems that deal with large numbers of interacting molecules, it has recently gained relevance within the Origins of Life community.[157] The main subjects studied within this framework are Dynamic Combinatorial Libraries (DCLs)[158,159], autocatalysis[160], oscillating reactions and self-assembling systems. A review on the subject by Ruiz-Mirazo *et al.* has been recently published.[94]

However, it was not until a recent article by John Sutherland and co-workers was published that the 'systems' approach was considered in the bottom-up synthesis of life's precursors.[161]

Sutherland describes a series of interconnected reactions in which the main precursors of life's subsystems (nucleotides, amino acids and lipids) can be synthesised from very simple starting materials: HCN and H_2S.[69] This work has changed the view in which classical prebiotic experiments are performed: instead of focusing on the individual synthesis of nucleobases or amino acids, this work explores all the different prebiotically relevant products that can emerge from two 3-atom molecules.

As it is typical in this area of research, many arguments were raised against how this work could fit into the 'systems' rationale due to the fact that the products obtained are often purified in each synthetic step and a high level of human intervention is required.

Regardless of the debate on what should be considered systems chemistry, it has become clear that heterogeneous mixtures containing several precursors could lead to the formation of more interesting and complex molecules due to their chemical and physical interactions, such as: interconversion processes, polymerisation reactions or supramolecular aggregation.

Previous experiments have already shown synergistic effects when different chemistries are combined: Weber demonstrated that ammonia and amines (including amino acids) catalysed the formation of sugars from formaldehyde and glycolaldehyde[162]. Also, some of the compounds produced in this reaction are catalysts of the same reaction, making it autocatalytic.

The role of compartmentalisation in the synthesis of simple precursors has also been explored: Adamala and Szostak have recently shown that vesicles enhance the formation of peptide bonds in the presence of a dipeptide catalyst.[163]

Also, Davis showed that the encapsulation of the formose reaction within vesicles led to changes in the final product distribution[164].

These examples illustrate how interactions between different types of precursors contribute to changes in the system that would not occur if pure starting materials were used instead. In light of these results, and taking into account the problems encountered when attempting to synthesise nucleic acids from very pure precursors, it is clear that a systems approach

might provide new insights into the transition from simple organic molecules to self-organised systems capable of Darwinian evolution.

1.5 Future outlook

Most experiments performed in prebiotic chemistry have usually been optimised to produce very specific products under very specific conditions. Once the optimal conditions are found, further efforts are made to fit the results into a particular narrative (i.e. RNA-first scenarios). This has resulted in the lack of compatibility between different systems, and the inability to produce complex enough polymers with catalytic or self-replicating properties. In this sense, changing the way experiments are performed could be very useful. Using high-throughput screening and combinatorial approaches will help us to explore a vast number of parameters that could significantly contribute to the discovery of new, otherwise unidentified chemical pathways. This would require understanding of chemical reactions and the effects of different physical parameters. Thus, chemical engineering can play a central role in advancing the field.

In this line of this thought, Dave Deamer has envisaged an automated platform which will eventually support the formation of proto-life forms from simple chemical building blocks. In his more recent book *'First Life'*[165], he describes how this automated programmable platform ('The Genesis Engine') would promote the formation of biopolymers from a heterogeneous monomer mixture. This would be accomplished by studying several parameters simultaneously (pH, temperature, etc.), with several analytical techniques (HPLC, mass spectrometry, capillary electrophoresis and electron microscopy). This model reactor would require a long time to set-up and the costs would be very high, however, the development of 3D-printing technologies in chemistry and the facile programming of affordable and accessible electronics (*i.e.* Arduino) have recently contributed to the development of cheap, custom-made robotic platforms capable of performing a vast number of experiments.

In a recent publication from our research group[166], we described the assembly of a custom-made 3D printed robotic platform that uses artificial evolution to select for desired behaviours in chemical systems. In this case, the macroscopic behaviour of oil droplets (i.e. droplet motion, division) was studied. The robotic platform was capable of: a) generating different oil formulations using syringe pumps, b) placing a few oil droplets in a

petri-dish containing aqueous phase and surfactant, c) recording a video of the droplet's macroscopic behaviour, d) processing and analysing the data and e) cleaning itself to start a new cycle. This process was completely automated and it was capable to run autonomously for an indefinite number if cycles.

Different ratios of four different oils were studied in this experiment. In order to explore all the possible unique combinations between these four oils, 17 million experiments would have been required. This would 575 years to be completed. However, a unique approach was taken to overcome this problem: a genetic algorithm explored different droplet behaviours within the vast parameter space instead of screening every possible combination. A genetic algorithm is an optimisation method inspired by natural selection. The algorithm generates a series of random experiments (individuals) with a set of different parameters for each individual experiment (genotype), which are then evaluated according to a fitness function. Then, the 'fittest' individuals are used to generate new ones by means of mutations and crossover. As this process is repeated over successive generations, the fitness value of the population is increased.

The use of custom-made robotic platforms constitutes a cheap and efficient alternative to standard automated platforms, making it accessible and affordable to the majority of researchers.

Alternatively, the use of microfluidic technology has also been suggested as a promising tool to be used in Origins of Life research. A summary of the state-of-the-art and applications of microfluidics is described in the following section.

1.6 Microfluidics

Microfluidics describes the study and development of systems that control small volumes of fluids (typically on the picolitre to nanolitre scale), have enhanced analytical performance, low unit cost, and the ability to access a large number of individual experiments per unit time.

Unfortunately, an efficient manipulation of fluids on such small scale is hard to achieve. Flow-based microfluidics, which are used in applications such as centrifugal fluidics, perfusion platforms or large scale integration, often become impractical due to Taylor dispersion, solute surface interactions, cross-contamination and the need for substantial volumes of reagents.[167]

An alternative method to overcome these limitations can be found in reagent compartmentalisation within picolitre–nanolitre sized droplets that reside within a continuous and immiscible fluid.

The field of droplet microfluidics deals with generation and manipulation of droplets with dimensions in the range of femtoliter to nanoliter in volume.

Lab-on-a-chip platforms have become a useful and extensively studied tool in chemical and biological fields, such as microreactors[168,169], drug delivery systems[170–172], microcapsules and microparticles[173–176], chemical analysis[177,178] and cell research.[179–181]

More specifically, such systems have found application in chemical and biochemical screening[182], protein crystallisation[183] and enzymatic kinetic assays[184], and will have a significant impact on other fields such as emulsion based PCR[185], DNA sequencing[186], directed evolution of proteins[187–190] and cell based binding and sorting assays.[191–193]

Micro droplets can be generated by combining two immiscible phases (typically water and oil). The formation of droplets is a spontaneous process and is a result of shear force and interfacial tension at the liquid–liquid interface.

Therefore, manipulation of these droplets with high precision enhanced the realisation of diverse operations within a device. Five approaches to manipulation of droplets can be found in the literature based on the sources of the force involved: hydrodynamic stress, electrohydrodynamics, thermocapillary, magnetism and acoustics. In the process of manipulating droplets, these forces have their own characteristics, in terms of formation, fusion, fission, mixing, sorting and transport of droplets (Fig. 12).[194]

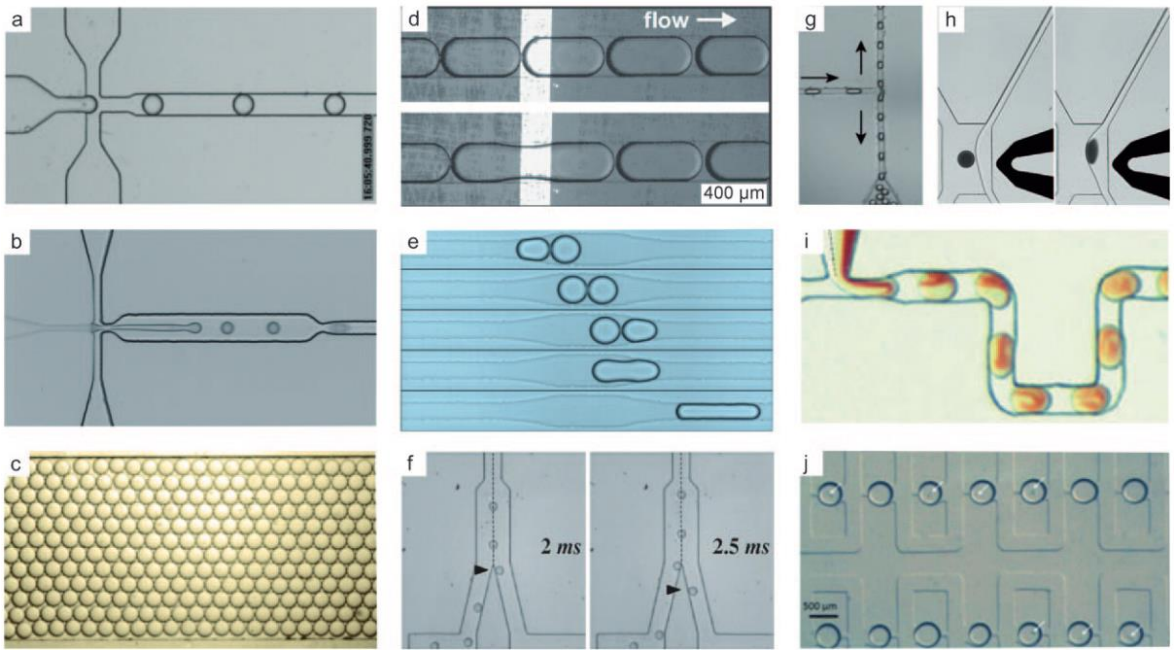


Figure 12**:** An overview of commonly used droplet-manipulation modules. a) Droplet formation in a flow-focusing device. b) Droplet formation from jetting in a flow-focusing device. c) Delay line/storage area. d) Droplet fusion by electrocoalescence. e) Passive, channel-geometry-mediated droplet fusion. f) Electrosorting of droplets. g) Channel-geometry-mediated droplet splitting. h) Electric-field induced emulsion separation. i) Mixing the contents of droplets in a winding channel. j) Droplet storage. Adapted with permission from John Wiley and Sons, Copyright © 2010 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.[195]

Various methods have been described for fabricating microfluidic devices, such as glass etching and soft lithography in poly(dimethylsiloxane) (PDMS)[196], which allow precise control of the microchannel geometry and can produce large quantities of complex devices. However, their manufacturing processes are complex, and require the use of expensive equipment.

During the past two decades, microfluidics has evolved from a technology built by a small community of academics into a technology with serious commercial potential, exploited by companies providing instruments for DNA, RNA and protein applications.[197]

Whilst microfluidic technology has not been extensively used in Origins of Life research, it has a promising future. A few recent examples in the literature describe how liposomes and water-in-oil droplets generated in microfluidic devices can be employed as artificial cell analogues. The aim of this type of research is to produce micrometre-sized structures that mimic many of the morphological characteristics of a living cell. A thorough review on the recent advances achieved in this research area has been published both by Andrew DeMello[198] and Stephen Mann.[199]

2 Aims

Many aspects related to how living systems could arise from simple chemicals still remain unanswered. For the last 50 years, traditional synthetic approaches have been used to investigate the formation of precursors that constitute living systems; however, the recent advances on laboratory automation and high-throughput analysis can change the way these investigations are performed, leading to the exploration of a much larger number of possible reactions under different conditions.

Following this, the primary aim of this research is to design and assemble automated systems capable of autonomously performing different unit operations to explore a prebiotic chemical space and its pathways to the formation of polymers relevant to the origin of life.

Initially, we will focus on the development of microfluidic technology that allows us to perform a vast number of experiments by encapsulation of reagents in micro droplets. This technology would also allow us to selectively mix the contents encapsulated in the droplets and apply a selection pressure against a fitness function, so we could use them as protocell analogues capable of undergoing Darwinian evolution.

In addition to the above, we want to use these automated platforms to investigate the formation of the polymers that constitute all living systems, in particular, the synthesis of long polypeptides. By using this approach, we aim to understand the mechanisms that govern the formation of peptide bonds in a fluctuating environment by a continuous high-throughput experiment that explores a vast number of physico-chemical conditions.

One of the major challenges that prebiotic chemists are still facing is the successful synthesis of peptide and RNA polymers of sufficient length that are able to fold into active structures that show catalytic properties or self-replicating capabilities. Therefore, we aim to control the synthesis of long heteropeptides mixtures that would ultimately possess the properties described above.

3 Results and discussion

3.1 Microfluidic platform for chemical evolution

The recent findings made in the Cronin Group on the assembly of inorganic molecules capable of showing life-like properties (self-replication from purely inorganic molecules[200], inorganic membrane/tubes formation[201]) have shed light on the possibility of alternative chemistry-based life systems.

However, since all life forms ever known are based on the same chemistry (DNA, RNA, Proteins and compartments), one must question what requirements alternative based-chemistries should possess in order to be considered a living system.

Whilst many definitions have been proposed on what should be considered as 'alive', it is widely accepted that a living system requires a minimum set of essential properties:[202,203]

- A population of individuals.
- Inheritance from one generation to the next. (Accrue self-derived bits of information over generations).
- Introduction of variations (mutation) in successive generations.
- A selection process against a fitness function.
- Self-replication and surviving of mutation over time

We can imagine a scenario where simple molecules within a mixture start to react with each other. Several newly formed molecules will arise, of which some would possess self-replicating capabilities. For replication to be successful, the self-replicating molecules must produce additional copies of themselves faster than existing copies become degraded. This can only be achieved in a pure system that contains the replicator and its corresponding building blocks. However, this would not be accomplished within a complex chemical mixture due to interaction and competition with other molecules.
This phenomenon could be avoided by isolating the replicator molecules and their building blocks within a compartment. In this way, new molecules will arise due to the accumulated mutations during the replication process, leading to the formation of new species with improved self-replicating properties.

However, in order to ensure their survival, these new molecules would have to transfer their information to the next generation via the division of the compartment where they reside into two smaller compartments. Upon iteration of these coupled processes (replication within self-dividing compartments, see Fig. 13), a competition between compartments will start for survival (the compartments containing the most efficient replicators will consume the available building blocks more quickly). After successive generations, the complexity and efficiency of the new species residing within the compartments would be greater than the initial ones. At this point, we could consider that this hypothetical chemical system fulfils all the requirements to be considered alive.

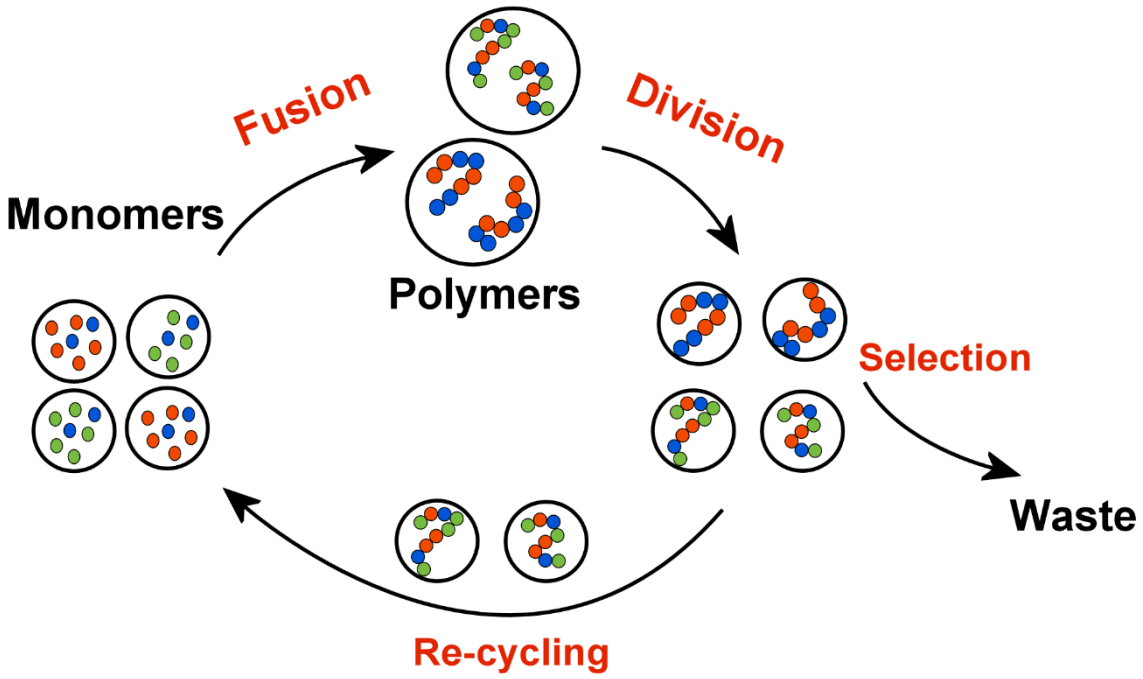


Figure 13**:** Schematic illustration of an alternative pathway to create a 'life-like' system from a bottom-up approach.

As mentioned in the Introduction, the recent development of droplet-based microfluidic technology has allowed the compartmentalisation of reactions at the micrometre scale, as well as the possibility of performing a vast number of individual experiments. However, how would microfluidic technology assist the formation of a minimal living system?

First, compartmentalisation within micro-droplets would avoid the problems associated with cross-reactions within complex mixtures. Then, the encapsulated contents within the droplets could be selectively mixed by using microfluidic manipulation to promote the formation of new species. The droplets would then be divided in order to transfer the newly formed species into two progenies. Finally, a selection process would be applied to those droplets containing molecules with improved replicating capabilities to ensure their survival within the system. By repeating these unit operations within a continuous loop, the encapsulated molecules will accrue new bits of information over successive generations, ultimately leading towards a system with enough complexity to show autonomous life-like properties.

In this chapter, the attention is not placed on the study of alternative chemistries with endowed life-like properties. Instead, the focus is placed on the development of an automated device capable of assisting the different unit operations required for molecules to undergo evolution in order to ultimately acquire properties of a life-like system.

Prior to the beginning of this project, the Cronin Group had not performed any microfluidics research, and hence, did not have the equipment and technology to do so. The first section of this chapter reflects the work I undertook on the establishment of a Microfluidics Lab, and all the further development and optimisation on fabrication techniques and materials. Then, more attention is placed on the fabrication of PDMS devices for droplet-based microfluidics. Finally, the design and testing of the different modules for a microfluidic device capable of assisting chemical evolution are described.

3.1.1 Microfluidics and soft lithographic techniques

Microfluidic device fabrication is mainly based on soft-lithographic methods.
Soft lithography is a collection of techniques based on rapid prototyping, printing, molding and embossing with an elastomeric stamp. This kind of fabrication combines high-resolution commercial printing, photolithography and soft lithography, and allows microfluidic systems to be designed and fabricated rapidly and inexpensively.

Rapid prototyping consists on designing a micro channel pattern with a scalable vector graphics editor. This design is then printed on an acetate sheet, which serves as a photomask in contact photolithography to produce a positive relief on a substrate.

Therefore, an elastomeric stamp with patterned relief structures on its surface with feature sizes ranging from 5 µm to 200µm is created by casting a liquid precursor against a master whose surface has been patterned with the complementary structures.

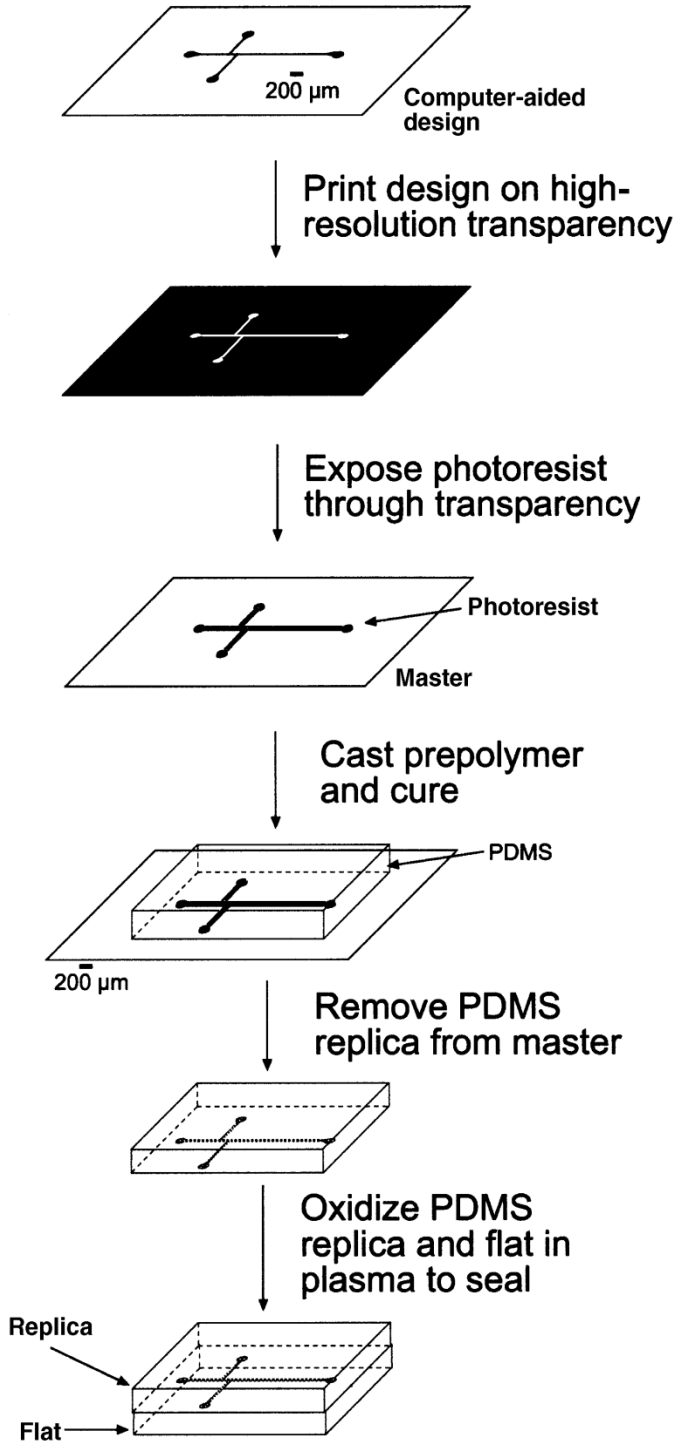


Figure 14: Fabrication scheme for polymeric microfluidic channels.[196]

As observed in Fig.14, a micro channel design file is printed onto a transparency photomask, which is then used in photolithography to generate a photoresist master. A pre-polymer of PDMS (Polydimethyilsiloxane) is cast onto the master, and the inverse replica is removed from the master and sealed onto a solid support (e.g., glass, silicon, or polymer).

3.1.2 Master fabrication

3.1.2.1 Photomask Design

A photomask is an opaque plate or film with transparent areas that allow light to shine through in a defined pattern. They are most commonly used in photolithography as templates to transfer patterns to wafers or other substrates in order to fabricate different sorts of devices.

Photomasks were designed on a vector drawing software. Once the photomasks were designed, they were sent to a photo service company to print them on a high resolution transparency film (Fig. 15).

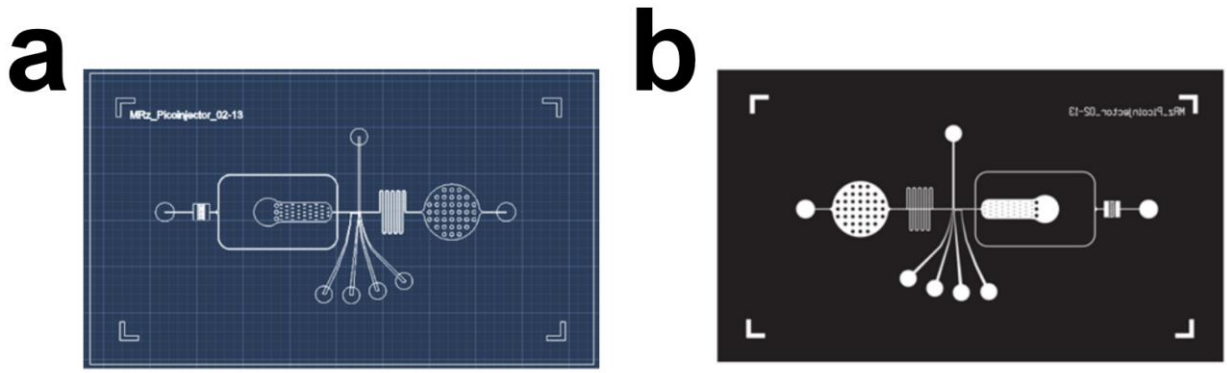


Figure 15: a) AutoCAD design of a microfluidic droplet pico-injector. B) A printed photo mask from the emulsion side.

3.1.2.2 Photoresist patterning

A photoresist master consists of a mixture of chemicals generally containing a film-forming polymer and one or more photosensitive compounds. When exposed to light of the proper wave-length, the photoresist undergoes photochemical reactions which alter the molecular structure of some of its components and change its solubility to form a patterned coating on a surface.

Photoresists are classified into two different groups: negative photoresists and positive photoresists.

Negative photoresists undergo photochemical reactions when exposed to a specific UV light wavelength. Consequently, the exposed area becomes insoluble to the photoresist developer and the photo mask design is patterned to the substrate. On the other hand, positive photoresists become soluble to the photoresist developer when exposed to UV light whilst the unexposed area remains insoluble. Therefore, each type of photoresist has its specific type of photo mask, as the exposed area undergoes a different reaction depending on its composition (Fig. 16).

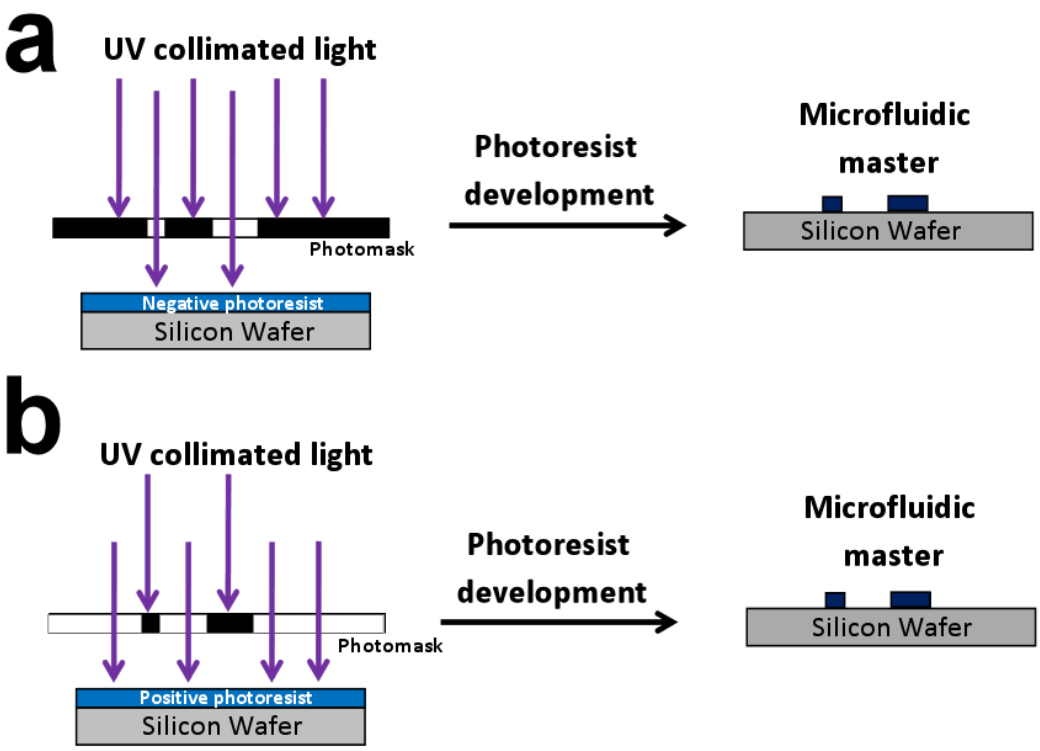


Figure 16: Schematic illustration of procedures for a) Negative Photoresist Patterning and b) Positive Photoresist Patterning.

In the following section, different results obtained for negative and positive photoresists are described.

3.1.2.3 Dry Film Resist (DFR) negative photoresist

Initially, we used ORDYL DRY FILM AM 100, which has a low cost and rapid development which makes it more suitable for faster microfluidic device fabrication. In these experiments, we used two different DFR: AM-140 (40µm thickness) and AM-175 (75µm thickness).
The first step requires careful control. Because the DFR film is laminated to a glass slide, the development process to achieve the desired pattern can become particularly tricky.

The whole procedure for DFR Template development is described in the Experimental section.

Several problems occurred when developing the photomask due to the thickness of the photoresist. AM-175 was used to fabricate micro channels between 100µm and 200µm. (Fig.17) However, when AM-175 was used to fabricate 50µm channels, at the last step of developing, the micro channels did not stick to the glass slide. Instead, AM-140 was tested for 50µm micro channel fabrication. By reducing the developing times, a 50µm micro channel photomask was achieved.

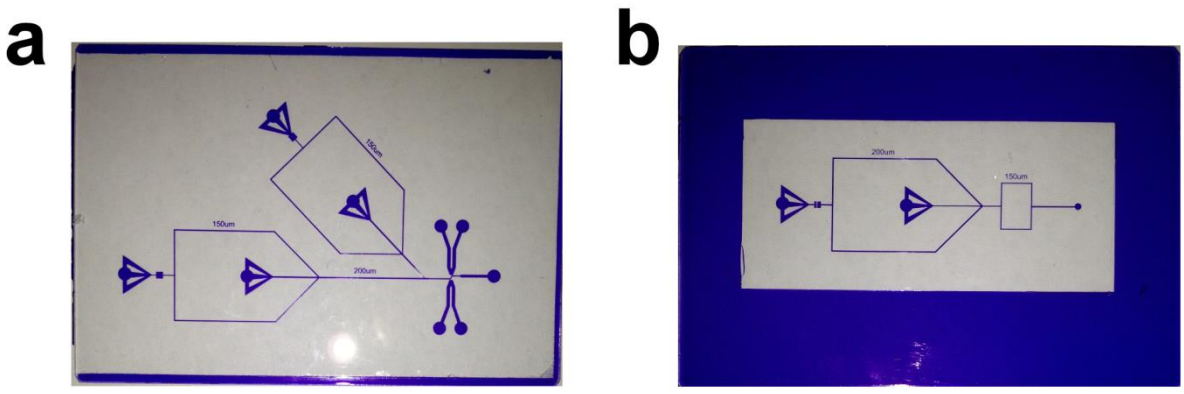


Figure 17: This photograph shows two different developed masks. (a) Droplet fuser mask and (100µm channels) (b) Droplet splitter mask (100µm channels).

3.1.2.4 SU-8 negative Photoresist

SU-8 is a high contrast, epoxy based photoresist designed for micromachining and other microelectronic applications, and it is widely used as a master in soft lithography. SU8 resists are highly functional, optically transparent and photo-imageable to near UV (365nm) radiation.
Depending on the resist thickness, different viscous resist products can be purchased. For this experimental work, we decided to use SU8-10 and SU8-25, which can be spun over a thickness ranging from 5 to 30 μm, and 15 to 40 μm respectively. The photoactive compound in SU-8 is dissolved in cyclopentanone, which improves the adhesion of the resist to the substrate. Upon exposure of SU-8, cross-linking of the resist proceeds in two steps: formation of a strong acid during the exposure process, followed by thermally driven epoxy cross-linking during the post exposure bake.

A standard process for fabricating a SU-8 master is:

1. Spin Coat
2. Pre-exposure bake
3. Exposure
4. Post-Exposure Bake
5. Development
6. Hard Bake (Optional)

These parameters have different values depending on the viscosity and the thickness desired, and can be checked in the protocol available from the company website. However, some steps should be optimised in relation to the experimental conditions. A detailed protocol for the fabrication of SU-8 masters is described in the experimental section.

In the initial experiments, several bubbles rose to the surface of the photoresist after the pre-exposure bake. This phenomenon could be explained by the accumulation of some dust on the resist surface that bursts during the baking steps. We partially solved this problem by moving all the equipment to a clean room and keeping all the materials dust-free. Nevertheless, a few micro bubbles rose to the surface of the resist in later experiments.

Dissolved air micro bubbles which are not easily observable when the resist is poured onto the substrate rose to the surface when the solvent was evaporated during the baking steps. A recommendation to overcome this problem is to heat the resist prior to spinning. Thus, the viscosity is lowered and bubbles float to the top.

After the resist has been spun onto the substrate, it must be soft baked to evaporate the solvent and densify the film. To improve the resist-to-substrate adhesion, it is highly recommended to ramp or step the soft bake temperature. Lower initial bake temperatures allow the solvent to evaporate out of the film at a more controlled rate. Finally, one of the crucial steps to success in this fabrication process is to control the exposure. Some exposure tests must be done during the initial experiments to optimize the most appropriate UV dose. Otherwise, adhesion failure, negative sidewalls or excessive cracking will appear as a result of an under cross-linking condition. The typical values for SU-8 10 (25 µm thick film) exposure energy are 200 mJ/cm^2.

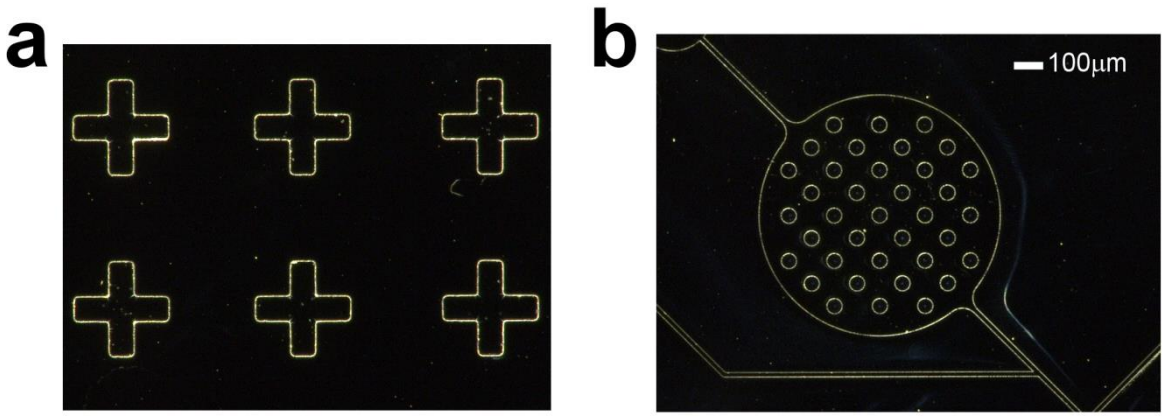


Figure 18: This photograph shows two different SU-8 master patterns. (a) Multilayer alignment marks. Some cracking is observed due to underexposure of the photoresist. (b) Droplet reservoir with some pillars.

3.1.2.5 SPR220-7.0 positive photoresist

SPR220-7.0 is a general-purpose positive photoresist designed to cover a wide range of film thicknesses between 1 µm and 30 µm in a single coat. SPR 220-7 is primarily composed of a diazoquinone ester and a phenolic Novolak resin as a photoactive compound. The lack of information from the supplier and the remarkable differences with the negative photoresists, resulted in several failures during the fabrication. SPR220 does not cross link when exposed, instead, the diazoquinone photoactive compound transforms into a carboxylic acid, accompanied by the release of nitrogen and the absorption of water. Therefore, the exposed area becomes alkaline soluble and can be easily removed by a suitable developer.

Unlike negative photoresists, positive photoresists do not cross link and start to soften between 100-130 °C. This phenomenon leads to a rounded cross-section of the resist pattern, which can be transferred into an elastomeric stamp by replica molding. In this experimental work, the purpose of using re-flowed photoresists was to fabricate multilayer microfluidic devices, where round-shaped channels are required for the actuation of micromechanical valves.

The standard process for fabricating a SPR220-7.0 master described in the official photoresist protocol is:

1. Spin Coat
2. Pre-exposure bake
3. Exposure (not precise)
4. Post-Exposure Bake
5. Development
6. Reflow (Not mentioned)

After repeating this protocol a few times, we realized that some crucial information was missing.

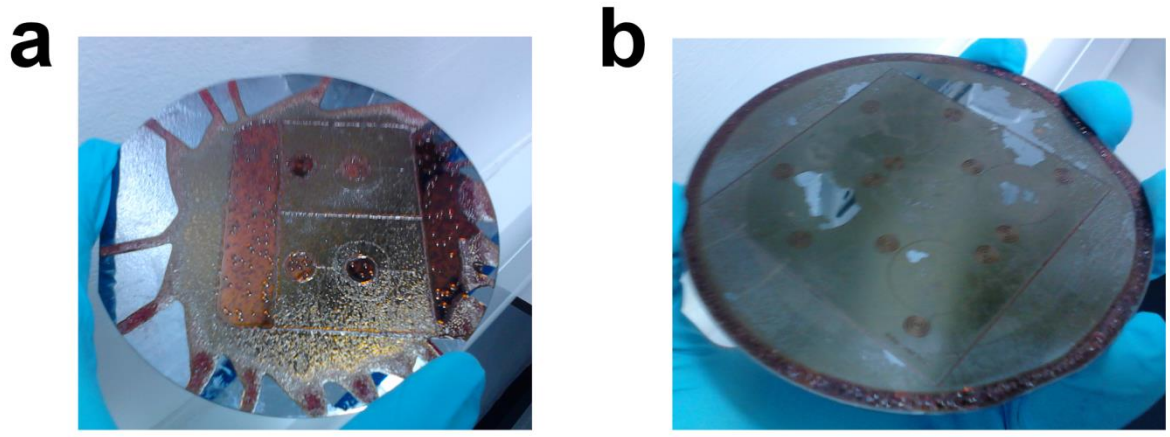


Figure 19: This photograph shows two different SPR220-7.0 master patterns. (a) Photoresist was not uniformly spun onto the substrate. Several bubbles are observed. (b) Edge beading was observed, and the photoresist was impossible to develop completely.

After asking for advice from three different experts in micro fabrication, we learnt some extra information that was not reflected in the protocol, which was critical for achieving a working procedure.

Firstly, it was clear that the photoresist was underexposed, as the exposed area remained partially undeveloped (Fig. 19). Also, during soft bake the bulk water concentration of the photoresist drops. However, some water content in the resist during exposure is required to allow a reasonable high-development rate. For this reason, a hold time of 24 hours is required between pre-exposure bake and exposure, to allow the photoresist to rehydrate.

The thickness of the resist plays an important role during the baking and exposure steps. SPR220-7.0 can be spun up to a 30 µm thick film in a single coat. Nevertheless, in the initial experiments we spun it two times to achieve a 40 µm thick film. When the thickness of the photoresist is increased, a higher exposure dose is required. In addition to this increase of the exposure dose, the consequent N_2 release of the photoactive compound may cause bubbles or cracks.

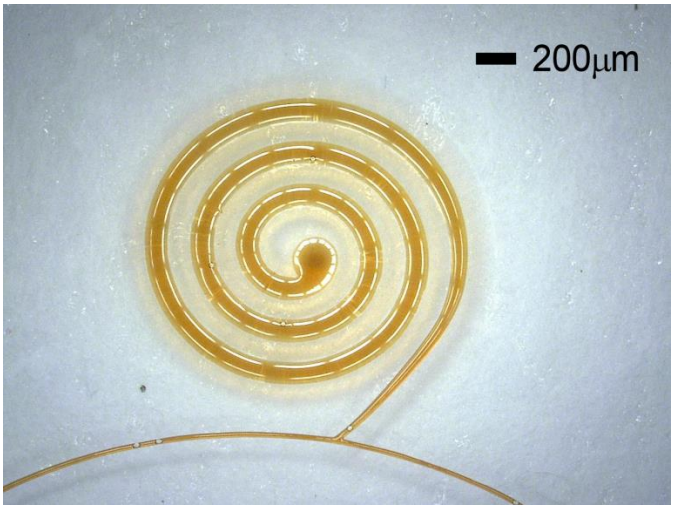


Figure 20: This photograph shows a SPR220-7.0 master on a glass substrate. (a) The photoresist showed some small bubbles that were collapsing the channels, thus making the device partially usable.

Furthermore, it is essential to have a precise control of the soft bake times and temperatures. Soft baking reduces the remaining solvent concentration by diffusion, which prevents foaming of the resists by N_2 created during exposure and improves resist adhesion to the substrate. We improved the quality of our masters by applying these changes (Fig. 20).

We also used HMDS (Hexamethyldisilazane) as an adhesion promoter prior to spinning, which partially improved the quality of the master.

Finally, several temperature tests were done to optimize the re-flow of the photoresist (Fig. 21). According to the literature, the softening temperature of SPR220-7.0 ranges between 110 °C and 115 °C, but different baking times are reported.

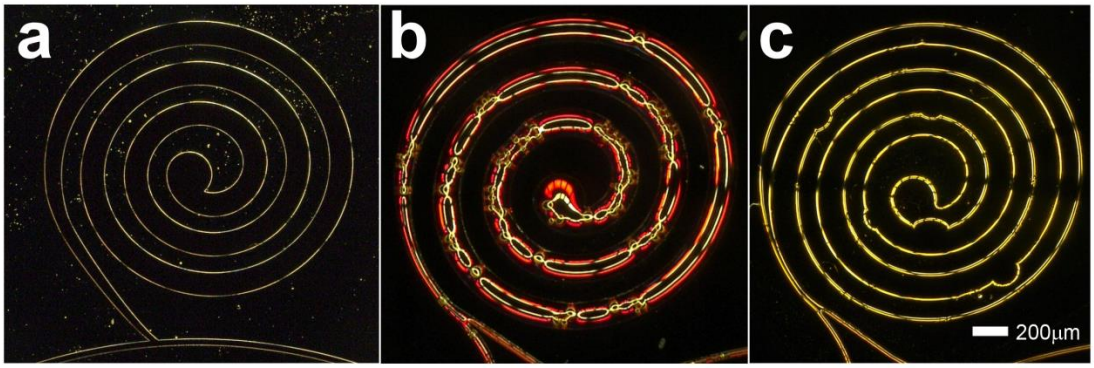


Figure 21: This photograph shows three SPR220-7.0 masters on a silicon wafer substrate. (a) A photoresist spiral pattern before being re-flowed. (b) A photoresist spiral pattern after being re-flowed at 120 °C for 6 min, collapsed by the appearance of several cracks and bubbles. (c) A photoresist spiral pattern after being re-flowed at 115 °C for 10 min (5 °C/min ramp).

After exploring several parameters to improve the quality of the masters, we found that achieving a highly reproducible protocol was hardly possible. As a consequence, we decided to purchase a widely used and commercially available positive photoresist: AZ-9260.

3.1.2.6 AZ-9260 Positive Photoresist

AZ-9260 is a thick film resist that provides high resolution with superior aspect ratios. As with other positive photoresists, AZ-9260 forms an indene carboxylic acid during exposure making it soluble in aqueous alkaline solution. The photoactive compound of AZ belongs to the group of diazonaphtho-quinone-sulphonates (DNQ). During exposure, the DNQ transforms into a carboxylic acid, accompanied by the release of nitrogen and the absorption of water. DNQ-sulfonate molecules are bonded to a *backbone*-molecule (a ketone) to improve the photoactive compound solubility in the resist. As SPR220-7.0, AZ-9260 requires a rehydration step between soft-bake and exposure.
Otherwise, if the resist lacks a minimum concentration of water between these steps, the ketone may perform various side reactions that can lead to an increase of the total dark erosion and low development rate.
Unlike SPR220-7.0, there is plenty of detailed information for each step of the fabrication process. Therefore, it just took us a few hours to fabricate a usable master with reproducible results (Fig. 22).

The standard process for fabricating an AZ-9260 master is:

1. Single Spin Coat
2. Pre-exposure bake (ramping the temperature 5 °C/min)
3. Let the resist rehydrate for at least 10 minutes.
4. Exposure
5. Post-Exposure Bake (ramping the temperature 5 °C/min)
6. Development
7. Reflow (ramping the temperature 5 °C/min)

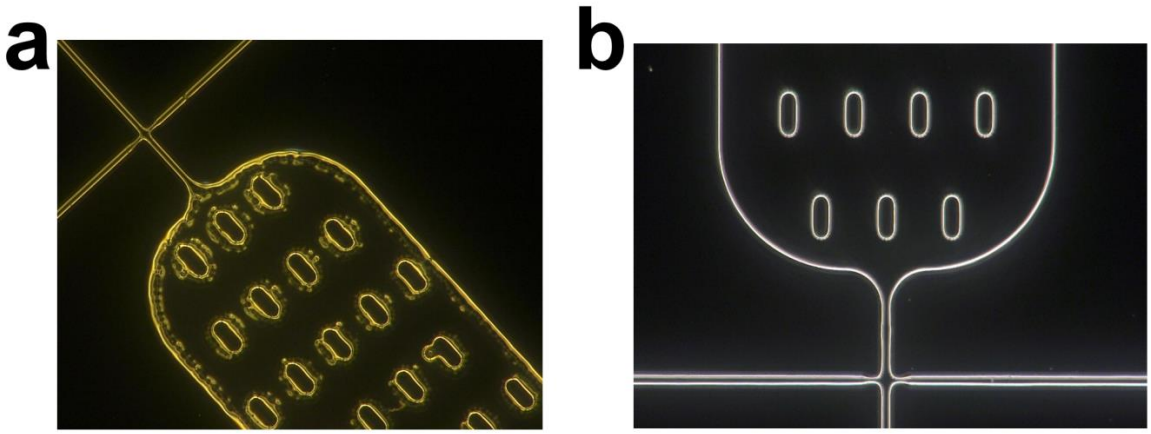


Figure 22: This photograph shows two AZ-9260 masters on a silicon wafer substrate. (a) A re-injection and flow-focusing geometry pattern after being re-flowed at 130 °C for 1 min. (b) Same design pattern after being re-flowed at 110 °C for 30 sec (ramping the temperature 5 °C/min) , showing perfectly rounded cross-section of the channels.

We finally decided to use SU8-10 and AZ-9260 as negative and positive photoresists, respectively. Once the baking and exposure parameters were optimised, master fabrication was highly reproducible and compatible for the fabrication of elastomeric microfluidic chips.

3.1.3 PDMS chip fabrication

PDMS (Polydimethysiloxane) is a flexible elastomeric polymer that is an excellent material for microfluidic device fabrication. One of the most common PDMS elastomers, Sylgard® 184 from Dow Corning® is used for microfluidic device fabrication. Sylgard is a two part resin system containing vinyl groups and hydrosiloxane groups. Mixing the two resin components together leads to a cross-linked network of dimethyl siloxane groups. Depending on the cross-linking ratio, the size of the string of monomer changes, thus leading to different levels of viscoelasticity.[204,205]

In microfluidics PDMS chip fabrication, the most common ratio for mixing the silicone elastomer and the curing agent is 10:1 (weight: weight). Once both parts are mixed for five minutes, the mixture is degassed by spinning in the centrifuge at 4.4 krpm for 5 minutes. Then, PDMS is poured onto the master with the channel side up, and cured at 65° for approximately two hours.

Once the material is hardened, because this material is flexible, it can be unmolded (peeled) from the master, leaving the master intact and ready to produce another device.[192]

3.1.3.1 Bonding PDMS

After cross-linking, PDMS becomes a hydrophobic elastomer. Activation of PDMS surface by oxygen plasma treatment is the most common method for bonding to another substrate, as it becomes hydrophilic for 30 minutes. When oxidized, PDMS produces silanol terminations on its surface. Then, it can be covalently bonded to an oxidized glass surface by the creation of a Si-O-Si bond.

Although oxygen plasma treatment bonding is the most effective method to create a non-reversible bond with a glass substrate, we initially tested two alternative methods:

The stamp and stick method consisted of depositing a sticky PDMS/TBA monolayer on a glass slide. Three drops of this mixture were poured on a clean glass slide, and spun for 30s at 1200 rpm. A PDMS slab was then placed on the glass slide, and soft baked at 65° until it was completely bonded.

Corona treatment for bonding PDMS is a simpler and cleaner method, compared to the ‘Stamp and Stick’ method. The high potential of the electrode ionizes surrounding air, creating a localized plasma or ‘‘corona.’’. Bonding is achieved by placing a clean glass slide and the PDMS sample bonding side up on a non-conducting surface, and exposing both surfaces to localized plasma. We used a Hand-held Corona treater, whose wire electrode was passed back and forth approximately 0.4 inch above each bonding surfaces. Treated surfaces were then pressed together and left undisturbed for at least one hour for bonding to take effect.

3.1.3.2 Device inputs

Prior to bonding, a biopsy punch of the correct diameter was used to drill the inlet ports. The punch was pushed through the PDMS until hitting the bottom, and the leftover PDMS section was discarded with a pair of forceps (Fig. 23).

To ensure the complete sealing of tubing connections in the chip, the size of the punched holes needs to be slightly smaller (~10%) than the outer diameter of the tubing. Otherwise, even if the inputs are sealed with another layer of PDMS, the right pressure will not be achieved when pumping, and undesired leaks will result.

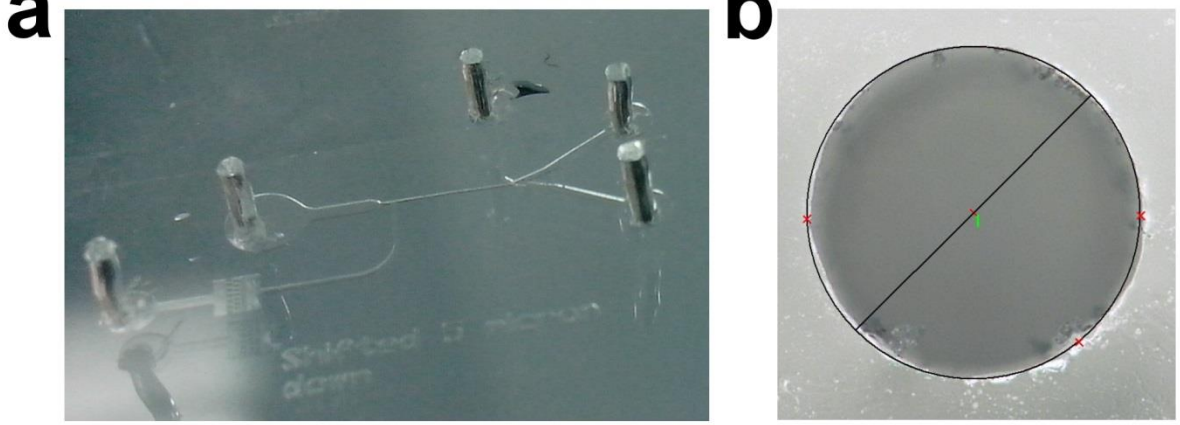


Figure 23: (a) A microfluidic droplet sorter with 5 fluidic inlet ports. (b) A 0.75 mm inlet port.

At this point, we had developed all the protocols and techniques to successfully fabricate PDMS devices for droplet microfluidics. The next step was to design and fabricate a microfluidic device capable of performing the required unit operations to assist the mixing, separation and selection of micro droplets within a continuous loop.

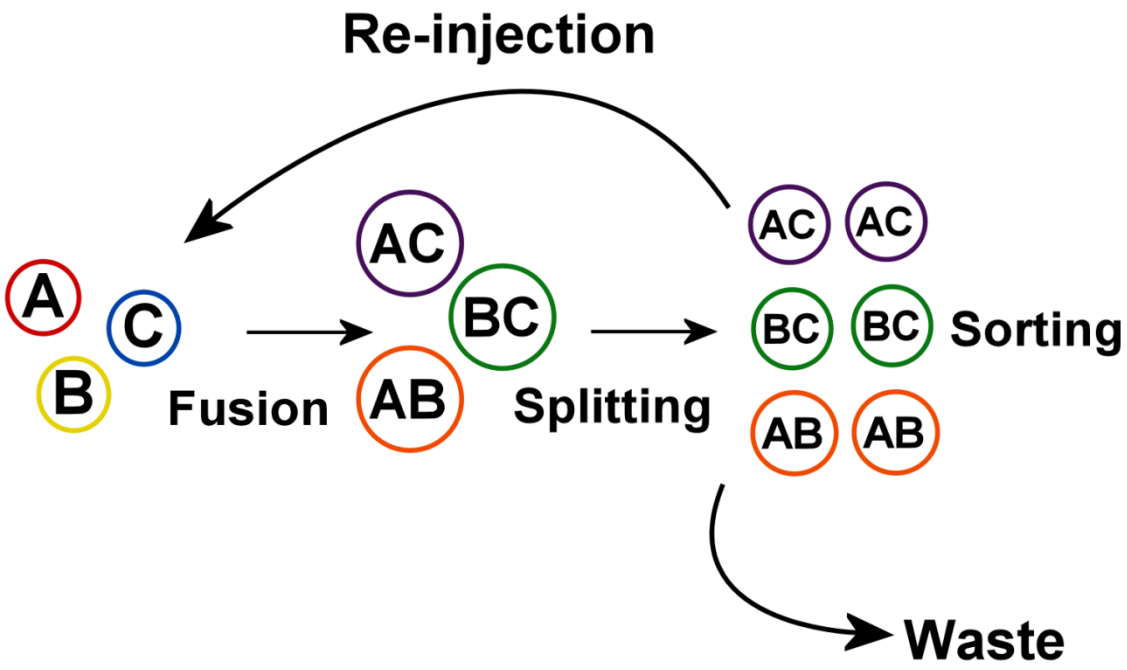


Figure 24: Schematic illustration of an embodied bottom-up chemical evolving platform where reagents are compartmentalized in picolitre droplets.

As depicted in Fig. 24, our ultimate aim was to design and fabricate a device containing:

1. A droplet generator.
2. A droplet re-injector.
3. A droplet splitter.
4. A droplet fuser
5. A system of micromechanical valves to transport the droplets within the different modules.
6. A droplet sorter.

The next section describes how individual modules were fabricated and tested to perform each one of these operations.

3.1.4 Droplet-based microfluidics

3.1.4.1 Microdroplet Composition: oil-in-water or water-in-oil emulsions

Droplet-based microfluidics can be employed to generate surfactant-stabilised liquid droplets by using two immiscible phases within determined micro channel geometry. Oil-in-water or water-in-oil monodisperse droplets are generated by using T-junctions or flow-focusing geometries.

The swelling of PDMS in organic solvents limits the performance of any oil-in-water droplet generation in a microfluidic chip. However, surface treatment of the microfluidic channels to make them hydrophilic has recently shown promising results for the creation of oil-in-water droplets.[206]

Over the past fifteen years, droplet-based microfluidic technology has expanded as a powerful tool to perform high-throughput analysis of water-in-oil micro droplets: Jean-Christophe Baret wrote a review about surfactants in droplet-based microfluidics[207], in which a list of commercial and home-made surfactants was reported for hydrocarbon and fluorinated oils. These oils, being both hydrophobic and lipophobic, are very poor solvents for organic molecules, and are thus especially well-suited for a separate compartmentalization of reagents within microdroplets.

Because of the low compatibility of PDMS with organic solvents, we decided to focus on the fabrication of microfluidic chips for the generation and further manipulation of water-in-oil droplets.

3.1.4.2 Flow rate VS pressure control

Two methods are widely used to drive the fluids from the pumps towards the device: by static pressure and by syringe pumps.
Syringe-pump systems change the pressure of the fluid to maintain a specified volume flow rate. In static pressure pumping, the fluids are placed in syringe tubes and the air pressure above the fluid is regulated itself.
We currently use a 6-channel high-pressure driven flow controller (MFCS-100, Fluigent).

3.1.4.3 Microdroplet Generation within a Microfluidic device

Droplet formation within the device micro channels is driven by hydrodynamic stress. Hydrodynamic stress is generated by an external mechanical force for driving two immiscible fluids. The most important parameters for droplet formation are: shear forces, interfacial tension between two immiscible phases and high resistance to continuous fluid flow. However, the parameter which determines the mechanism of droplet formation is the geometric design of the micro channels.

Concerning the geometry of the micro channels, two different shapes have been widely used for droplet formation: T-junction and Flow-Focusing device.

i) T-Junction: In this micro channel geometry, an aqueous phase is introduced perpendicularly, into the continuous oil phase. Droplet formation occurs due to shear force and surface tension competition. The size of droplets depends on the flow rates of the two immiscible phases, the viscosity, the interfacial tension, and the geometrical dimensions of the device.

ii) Flow focusing: In this case, micro channel geometry consists of three channels which converge into a main channel. Aqueous phase flow along the middle-channel, while two-side channels contain the oil phase. Both phases co-flow at the cross junction, where droplets are formed due to the forces generated from the continuous phase.

We focused on the fabrication of micro droplet generators with a flow-focusing geometry (see Figs. 25-27)

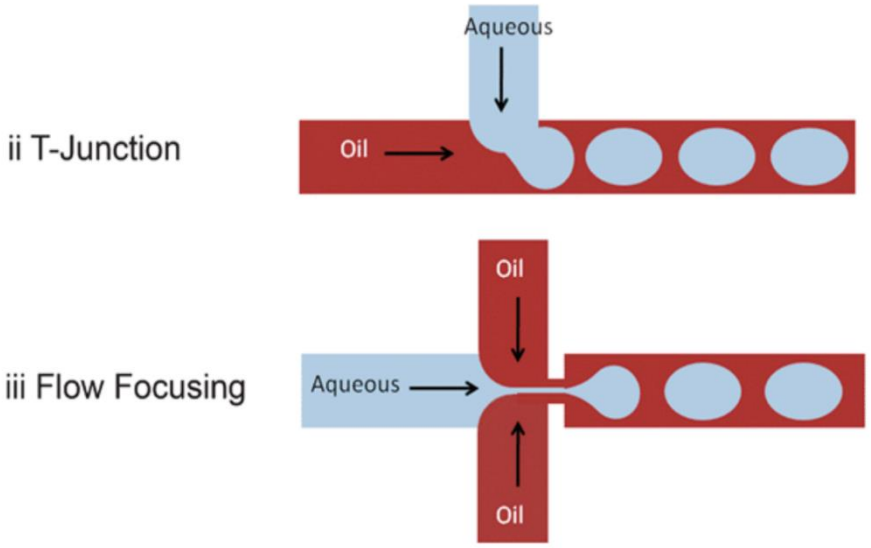


Figure 25: (a) T-Junction micro channel geometry. (b) Flow-focusing micro channel geometry. Both figures show two different geometries for aqueous droplet formation.[47]

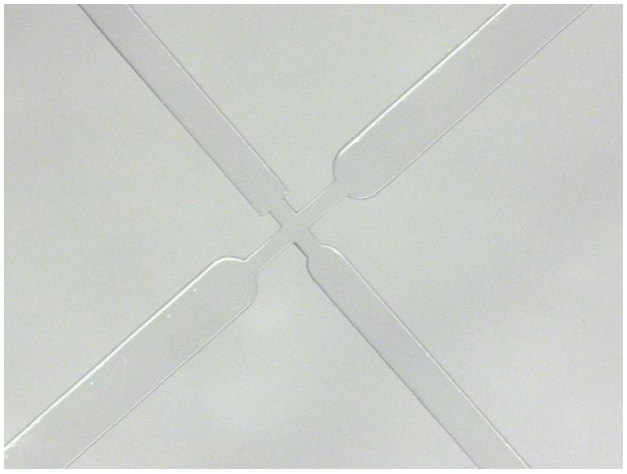


Figure 26: Flow-focusing droplet generator patterned on a PDMS chip.

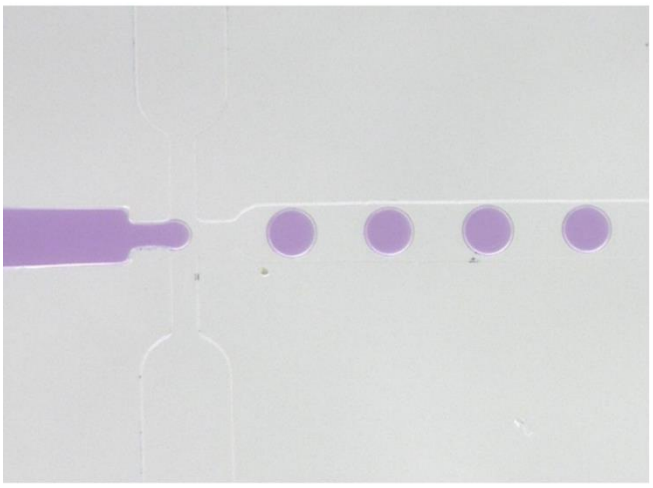


Figure 27: Uniform droplet generation within a flow-focusing geometry.

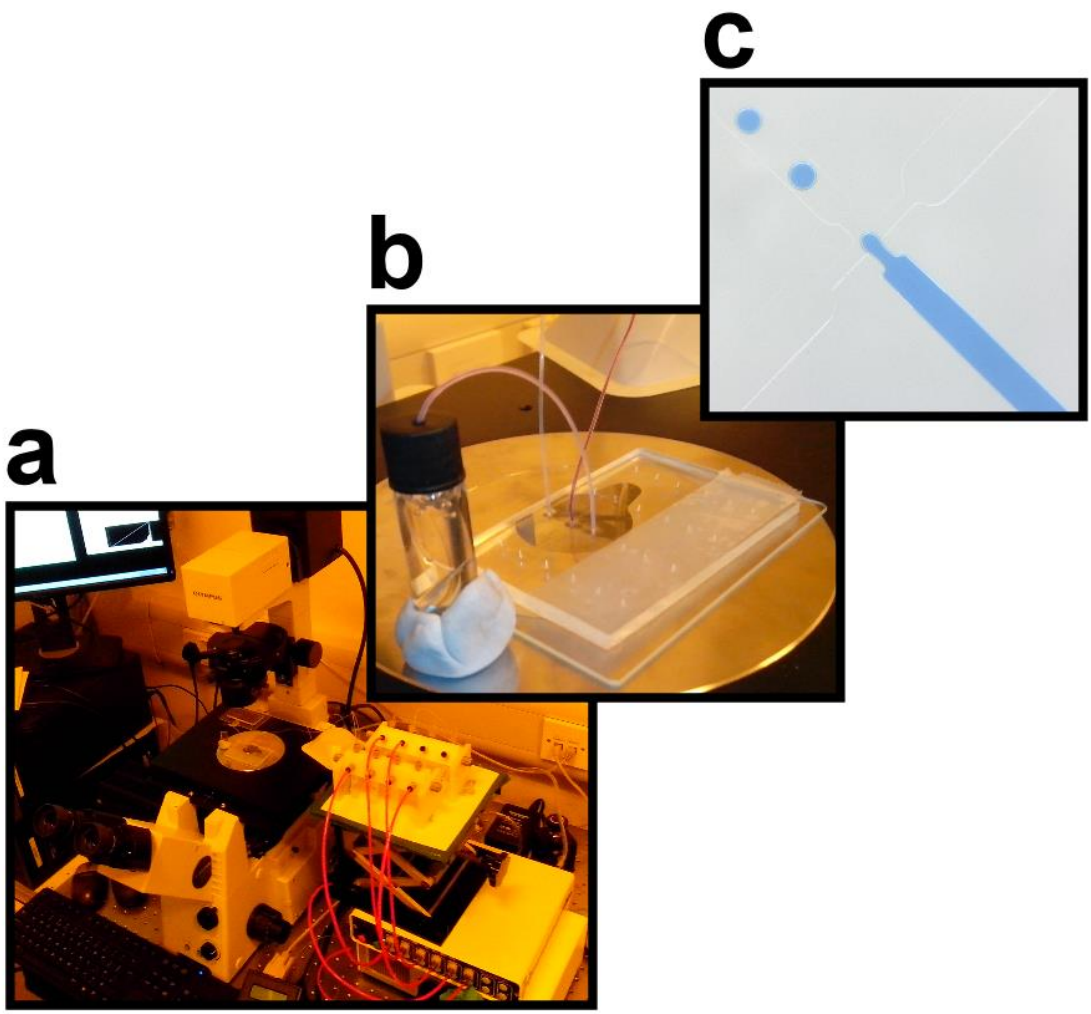


Figure 28: This photograph shows the entire set-up required for aqueous microdroplet generation. Droplet generation was recorded at 28 frames per second by an Olympus IX81 inverted microscope. (a) A 6-channel high-pressure driven flow controller (MFCS-100, Fluigent), connected to the microfluidic chip through 1/32'' PTFE tubing (Cole Parmer) (b) Eight parallel droplet generators in a microfluidic chip. The outlet port is connected to a collection vial for droplet storage. (c) Aqueous droplets generated at the flow-focusing junction.

In our droplet generation experiment (Fig. 28), high-pressure flow-driven pumping was used. Oil phase (2% EA surfactant in HFE-7500 Oil) and aqueous phase (distilled water and dye solutions) were loaded in 1 ml Fluigent vials, and placed in the Fluiwells (MFCS vials). PTFE tubes were filled with both oil and aqueous phase before connecting to the device. Pressure-rates for the aqueous phase (55mbar) and oil phase (50mbar) were set up. Different pressure rates were tested until achieving a uniform and continuous generation of a monodisperse population of microdroplets.

3.1.4.4 Droplet Re-injection

Droplet re-injection has been reported recently as a tool for droplet combinatorial synthesis; large libraries of compounds are compartmentalized in millions of micrometre-sized droplets that can be re-injected into another microfluidic chip to perform several operations such as fusing, splitting or sorting.[208] Also, this technique has been used for single cell-analysis and sorting. During re-injection, it is important to ensure that droplets pack together by removing most of the continuous-phase solution from the emulsion. Then, droplets are spaced by a continuous oil stream.

In the first attempt, droplets were generated at a fast rate and stored them in a 1 ml glass vial filled with the same oil phase. Droplets floated to the top of the oil phase when stored (Fig. 29).

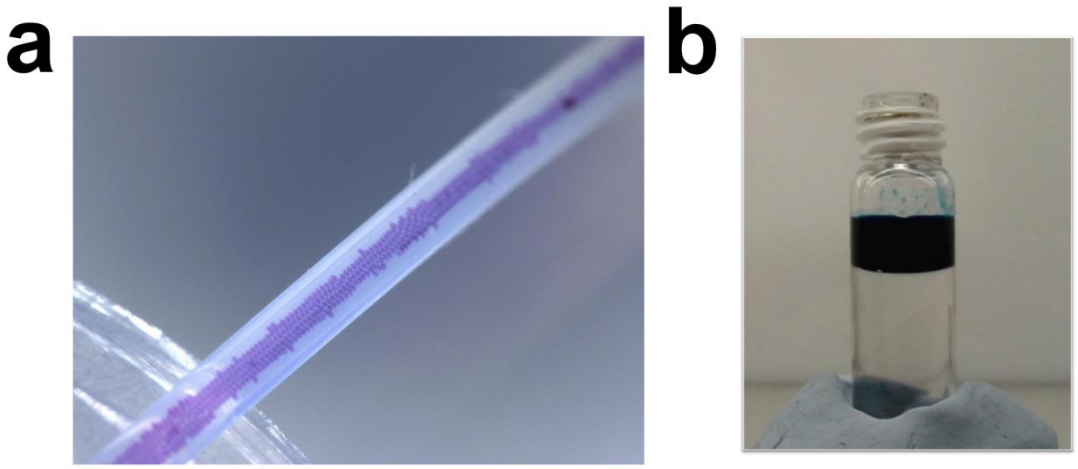


Figure 29: (a) Generated aqueous droplets transferred to a storage vial (50 µm in diameter) (b) Micro droplet aqueous emulsion sitting above the oil phase.

It is important to take into account that stored droplets are separated by a thin layer of surfactant. If the re-injection rate is too high or there are any obstacles through the tube, they can coalesce or split into smaller droplets, thus obtaining a highly polydisperse population of droplets. To avoid this phenomenon, it is important to remove all possible dust in the micro channels, as well as not to apply any external forces when the droplets flow through the tube.

In the first attempt, the emulsion was loaded into one of the Fluigent 1 ml vials, and the droplets were directly pumped into a microfluidic device. Unfortunately, some of the droplets coalesced when entering the chip. This could have been due to different factors, including:

- Any dust or leftover PDMS debris in the re-injection inlet.
- The tubing inserted into the inlet squeezes the droplets.
- The tubing should be inserted to one-third or one-half of the thickness of PDMS.
- Too-high re-injection flow-rates.

Some of the literature reported the use of specific inlet/outlet devices to perform the re-injection step:

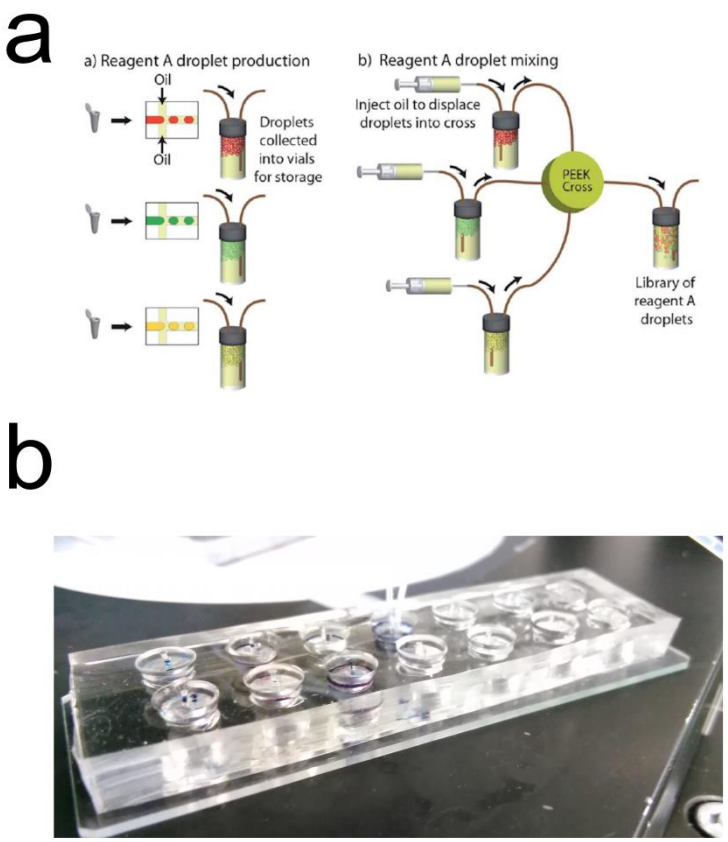


Figure 30: (a) Schematic illustration of commercial inlet/outlet devices for droplet re-injection (RainDance Technologies). (b) Home-made inlet-outlet device fabricated by bonding two parts of PDMS and punching two 0.75mm holes. 1/32'' OD PTFE tubing was used to connect the reservoir to the microfluidic chip and the pump.

By using our home-made inlet-outlet device (Fig. 30b), the emulsion was re-injected successfully into the microfluidic device, with a high degree of monodispersity (Fig. 31).

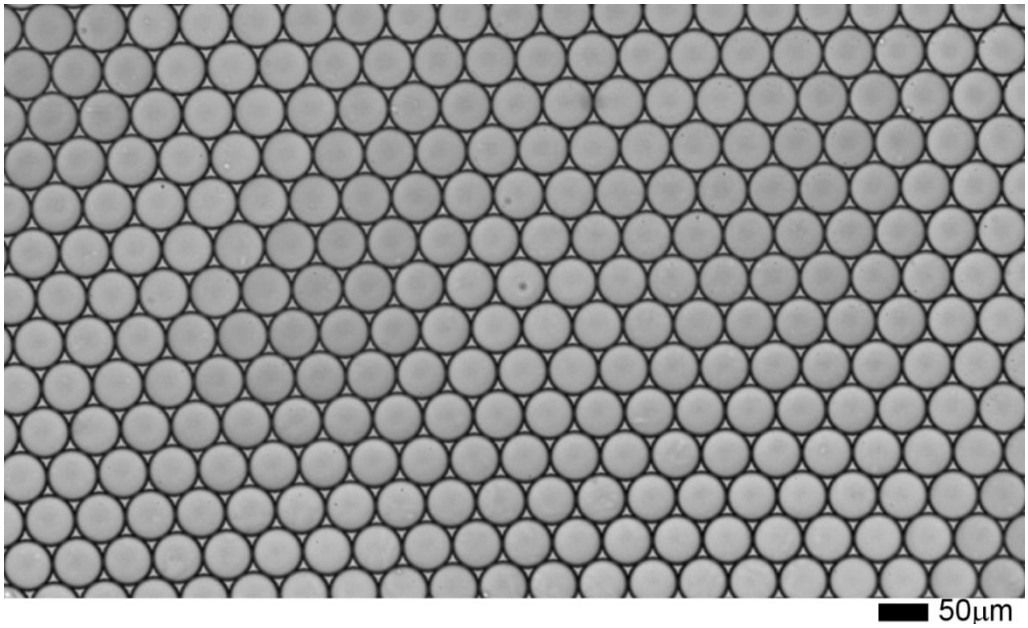


Figure 31: (a) 50 µm in diameter aqueous droplets being re-injected at a constant pressure rate of 50 mbar.

3.1.4.5 Droplet Splitter Device

Droplet fission can be carried out in a controlled manner by hydrodynamic stress and a bifurcating junction. Fission occurs when two-phase fluids flow toward a bifurcating junction, and the droplet is affected by the pressure and the shear strain arising from the flow. The relative size of the daughter droplets depend on the symmetry of the flow. If the flow is fully symmetric, the mother droplet will be split into two equal half volume droplets (Fig. 32).

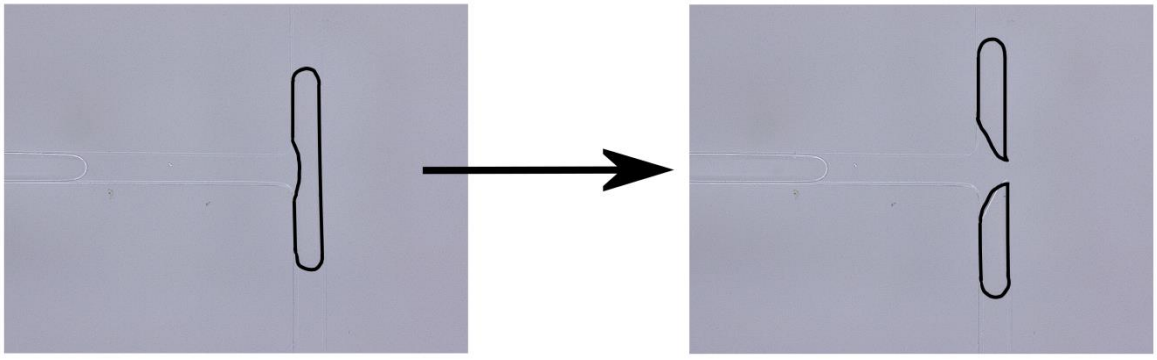


Figure 32: Aqueous droplets were generated in a 2% EA surfactant/HFE-7500 Oil solution within a flow focusing geometry. Droplets were then split in a symmetrical bifurcating junction. Daughter half volume droplets were obtained.

3.1.4.6 Droplet fuser device

In droplet microfluidics, fusing the droplets constitutes one of the most important unit operations for high-throughput experimentation. Whilst many different techniques have been described to fuse individual droplets, the most common way to induce the coalescence of droplets is by exposure to electric fields. This is normally achieved by flowing the droplets across a channel with electrodes on both sides, which polarise the droplets and induce their coalescence.

The fabrication of these devices primarily involves the insertion of microelectrodes (Fig.33). This is usually performed by flowing a low-melting temperature solder into a micrometre-sized channel and connecting it to a power supply through a copper wire. The overall fabrication of the device is difficult and time consuming.

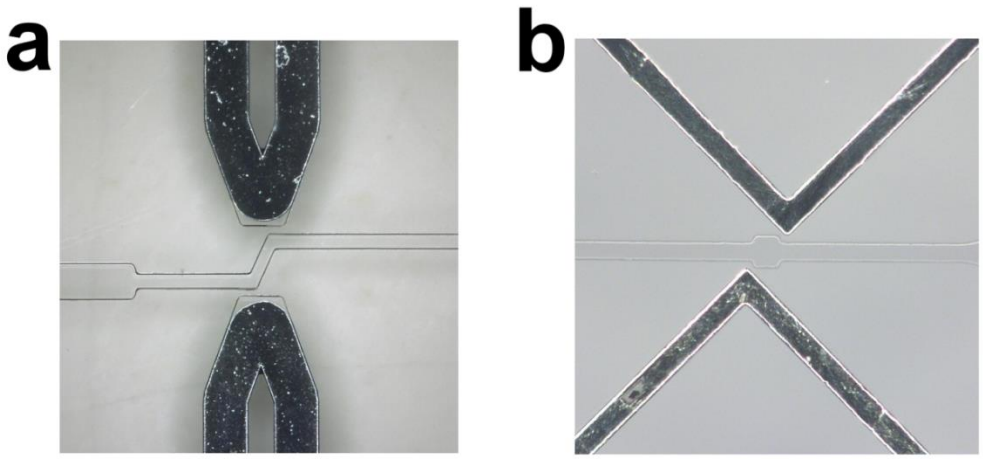


Figure 33: a), b) Inserted microelectrodes in a microfluidic device. Air bubbles are often trapped within the micro channel, thus preventing a proper electrical connection between the electrodes.

However, what constitutes a challenge is not the fabrication of the device but the synchronisation of droplet pairs before the fusion step. In order to achieve one-to-one droplet pairing, very precise control of the flow rate is required in each droplet generator module until the frequency of the two droplet trains is matched. This step is usually complicated and small fluctuations in the flow rate can lead to mismatch errors. However, different approaches have been applied to overcome this problem. One of them involves the fabrication of a railroad-like channel network that equilibrates the pressure between the channels where the droplets are flowing, ensuring up to 95% synchronisation efficiency.[209]

Therefore, we decided to design and fabricate a droplet fuser device containing a railroad-like channel network to facilitate the synchronisation of droplets before the fusion step (Fig. 34).

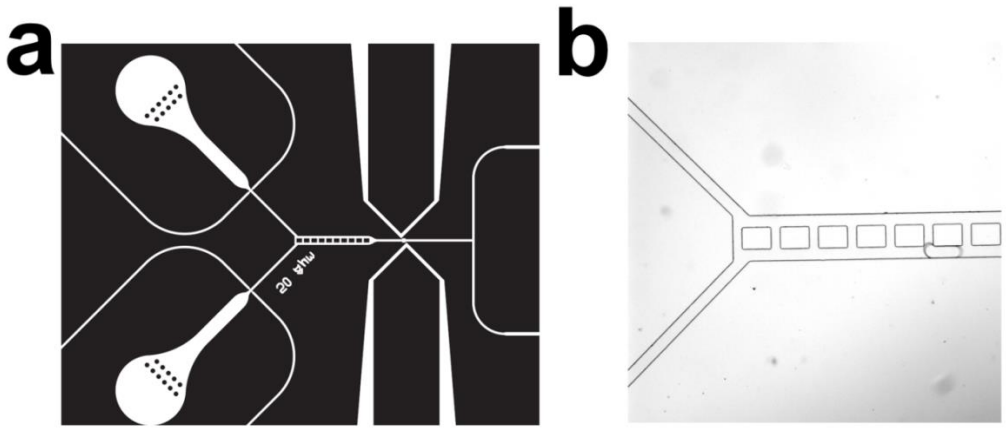


Figure 34: a) Expanded photo mask design of a droplet fuser-splitter device. Two droplet generators are connected to the railroad-like network through a Y-junction b) Railroad-like channel network for droplet synchronisation.

Whilst this design was created to facilitate the synchronisation of droplets, we could not observe in detail whether the pairing step was precise enough due to the limitations of the camera. This led us to recently purchase a high-speed camera capable of analysing at up to several KHz frame rate.

3.1.4.7 Multilayer Soft Lithography (MSL) and On-chip actuation

MSL is a technique that combines soft lithography with the capability to bond multiple patterned layers of elastomer. Using this technique, it is possible to create micromechanical valves by pumping pressurised air through one of the layers. Usually, MSL devices consist of two different elastomeric layers:

- Control layer: micro channels are pumped with high pressurised air.
- Flow Layer: micro channels are filled with fluids.

When a control channel and a flow channel cross, if the area of the intersection is large enough, a valve is created. The thin membrane separating the two channels deflects into the flow channel when the control channel is pressurised, creating a complete seal (Fig.35).

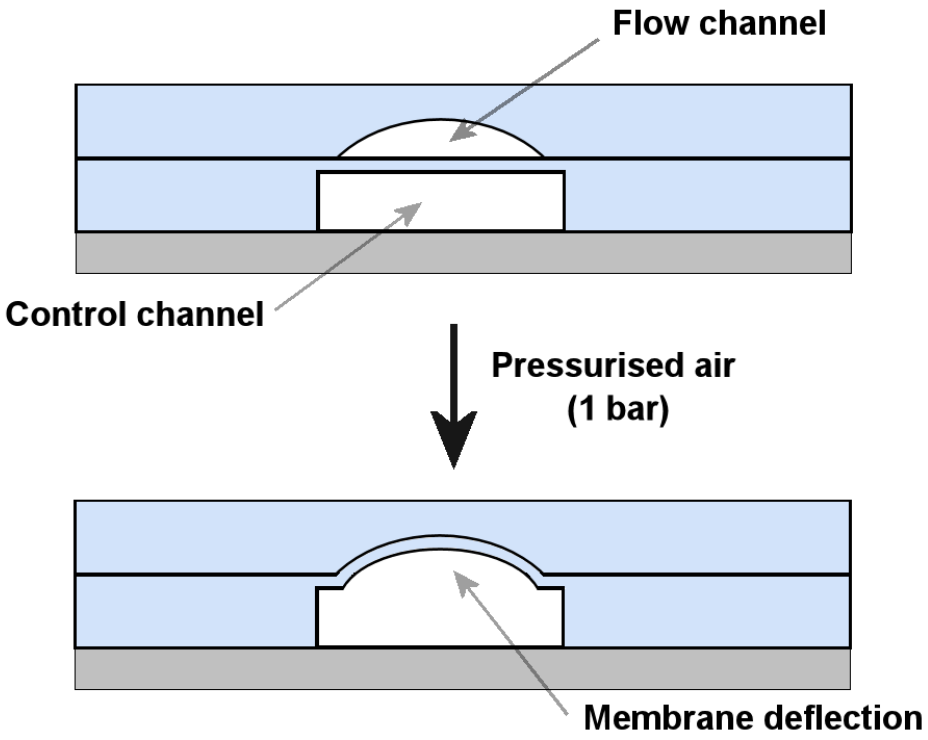


Figure 35: Schematic illustration of an actuated micromechanical valve by using Multilayer Soft Lithography.

An array of micromechanical valves actuated in a certain sequence can be used to perform different operations within the chip, such as: pumping, mixing and metering.

Also, the development of microfluidic chips with hundreds to thousands of integrated micromechanical valves enables hundreds of assays to be performed in parallel in an automated manner. This technology, commonly known as mLSI (microfluidic Large Scale Integration) was first invented and promoted by Stephen Quake. Several examples of this novel technology have shown promising results, like a microfluidic chip that can perform 400 PCR reactions using 41 pipetting steps versus the 1200 required on bench.[49]

However, only in rare cases the use of micromechanical valves for the manipulation of micro droplets has been described in the literature. Our aim is to fabricate microfluidic devices with a few integrated micromechanical valves that could perform some operations that cannot be performed in standard droplet based microfluidics such as storage or re-circulation.

The most important parameters to ensure the correct performance of a micromechanical valve are:

- Dimension of the valves in relation to the fluidic channels.
- Thickness of the membrane in-between the two layers.
- Pressure applied to deflect the thin membrane.

A more detailed procedure for fabricating multilayer microfluidic devices is described in the Experimental section.

Initially, we used SPR220-7.0 photoresist to achieve a 40 µm thick resist layer that could be transferred into a 40 µm tall flow channel (Fig. 36).

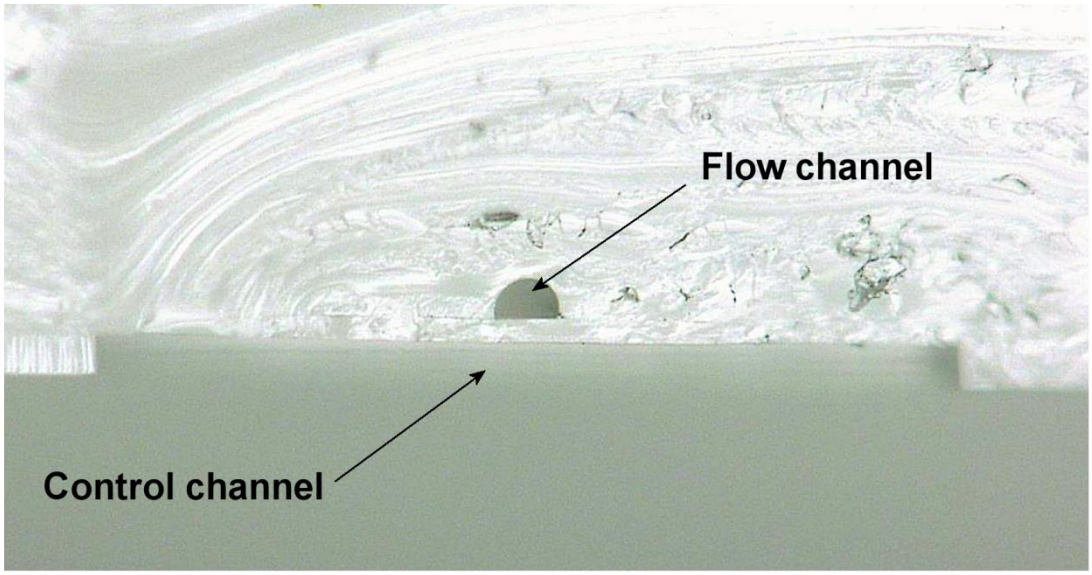


Figure 36: This photograph shows the cross section of a multilayer microfluidic device. The rounded flow channel was obtained by casting a 10:1 PDMS mixture onto a 40 µm thick SPR220-7.0 resist master. The control channel was obtained by casting 10:1 PDMS mixture onto a 30 µm thick SU8-25 resist master. Further measurements revealed that the membrane was 40 µm thick.

In order to test the reliability of the fabricated micromechanical valves, we designed a re-circulator microfluidic chip to move a population of droplets across three different reservoirs (Fig. 37).

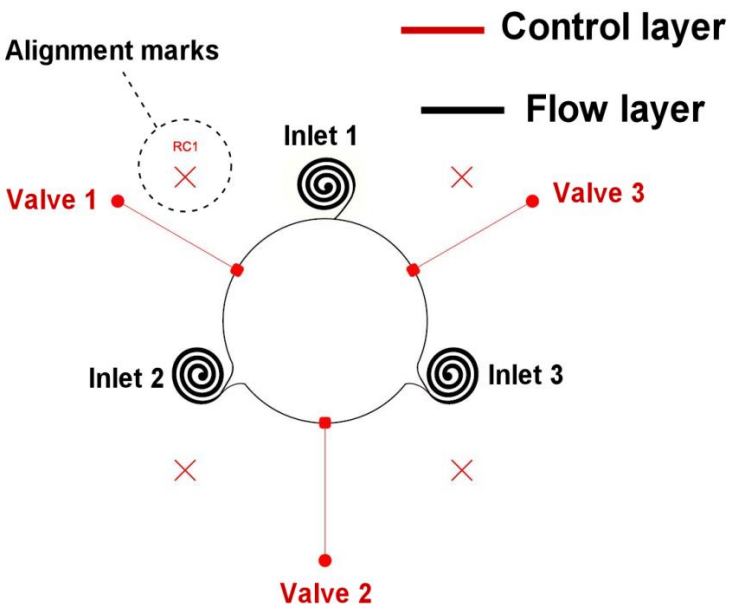


Figure 37 : (a) Schematic illustration of a droplet re-circulation device. Control and flow layers were bonded plasma treatment of both surfaces.

In this experiment, a previously generated population of droplets was re-injected into Inlet 1 whilst all the valves were closed. Then, to drive droplets anticlockwise along the different reservoirs, pneumatic valves and oil inlets were sequentially programmed to create a convective flow within the device.

Several attempts were made to transport a large population of droplets across the three reservoirs. Nevertheless, a problem occurred in relation to the deflection of the thin PDMS membrane between the two layers. Further search on specialised fabrication of MSL made us realise that the dimensions of the micromechanical valves had to be modified.

Most remarkably, we realised that the height of the flow channels could not exceed 15 µm. Otherwise, a higher pressure needed to be applied to the control channel in order to clog completely the flow channel by membrane deflection. In this case, this excess of pressure could lead to breakage of the thin PDMS membrane.

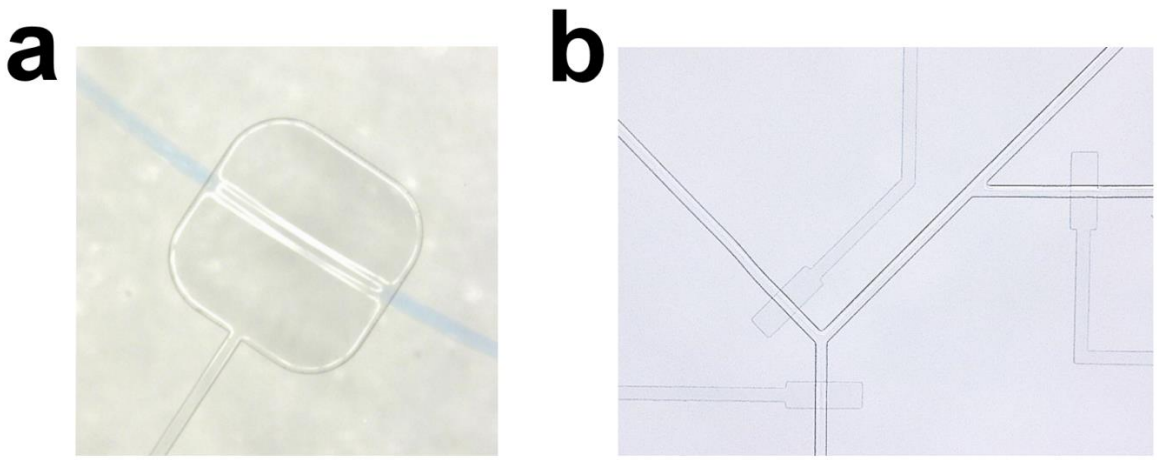


Figure 38: (a) This photograph shows an actuated micromechanical valve in a re-circulator device (The thickness of the flow layer was 40 µm). The valve was partially clogged and some material could flow through the two layers, b) a different multilayer device was fabricated for the manipulation and re-circulation of micro droplets, as observed, the dimensions of the valves were modified, as well as the thickness of the flow layer, thus resulting in a successful and fully working micromechanical valve actuation.

Further experiments showed promising results when the thickness of the flow layer was decreased to 10 µm, and the dimensions of the valves were slightly modified. In this case, we used AZ-9260 to create a 10 µm photoresist master of the flow layer. The control layer was obtained by casting 10:1 PDMS mixture onto a 20 µm thick SU8-10 photoresist master. See fig. 38 b).

3.1.4.8 Microfluidic Sorter

Microfluidic droplet sorting is widely used as a tool for selection and isolation of individual droplets within a population. There are two main different types of sorting: passive sorting and active sorting. Passive sorting is applied when the physical properties of the droplets are different (i.e. size or gravity), whilst active sorting involves electrical manipulation of individual droplets. Fluorescence activated droplet sorting (FADS) has become the most popular way of sorting droplets actively. Usually, individual droplets flowing through a straight channel are sorted within a Y-junction channel, which is designed with an asymmetry so the droplets would prefer to go towards one channel.

When the voltage at the electrodes is turned on, the electric field displaces the droplets towards the other channel. The droplets are sorted upon fluorescence. This is accomplished by shining a laser spot to each individual droplet before the Y-junction. The emission is recorded and processed by appropriate software, which then triggers the electrode actuation (Fig. 39).

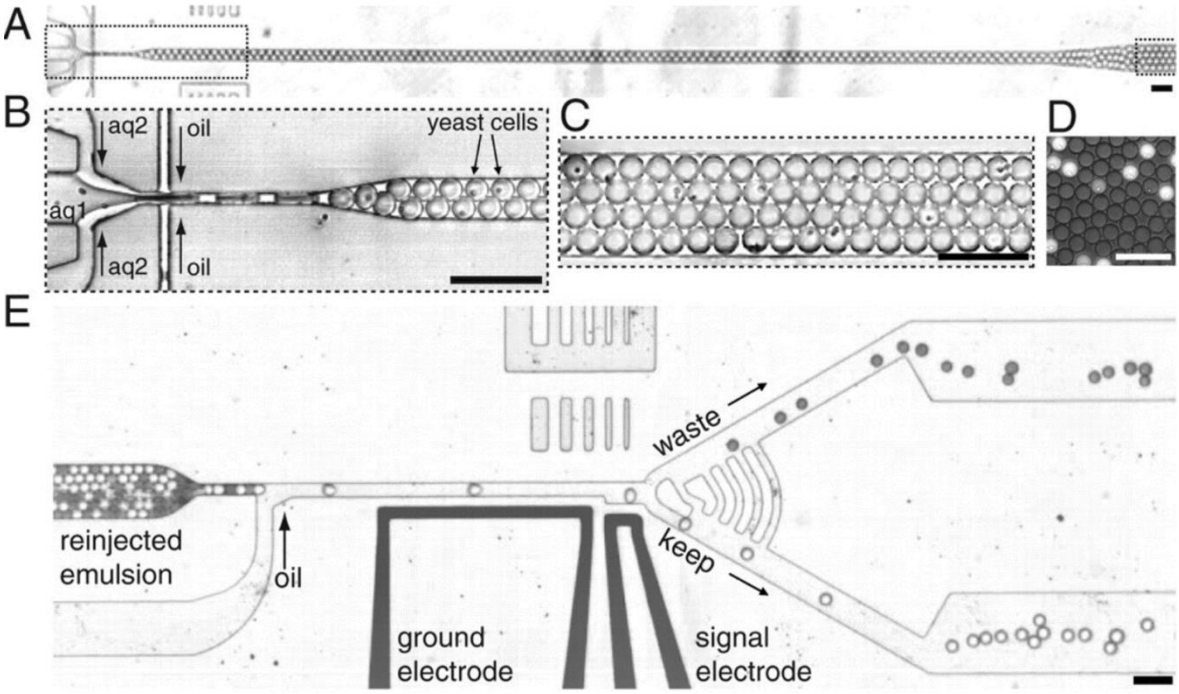


Figure 39. FADS device used for directed evolution of enzymes.[210]

This type of sorting has enabled researchers to overcome the problems associated with FACS (Fluorescent Activated Cell Sorting) by reducing the volume of the samples and their evaporation within the micro-titre plates. It also allows the screening of a much larger number of experiments, since every single micro droplet constitutes an individual assay. Reagents can be added to the droplets by fusion or pico-injection, thus increasing the versatility of the assays.

Whilst this technique offers reliable and high-throughput sorting, the device fabrication and operation requires much specialised expertise.

This is why we decided to explore the possibility of performing the sorting step with an alternative mechanism. Since we successfully optimised the fabrication of elastomeric micromechanical valves, we decided to investigate their application in droplet sorting.

3.1.4.9 Multilayer device for droplet sorting

In section 3.1.4.7, we described the fabrication of multilayer devices with integrated micro mechanical valves. We aimed to fabricate a device that could sort subpopulations of droplets by switching valves and controlling the hydrodynamic flow on the channels. However, only in rare cases has the use of micromechanical valves for the manipulation of micro droplets been described in the literature. The channel dimensions on multi-layer devices are usually smaller than the ones used in standard microfluidic chips due to the specific high-aspect ratio dimensions to ensure valve clogging.
As a first attempt, we designed and fabricated a multi-layer device with four micro mechanical valves to test their performance. The geometries of the micro channels were optimised to achieve a reliable valve actuation (Fig. 40).

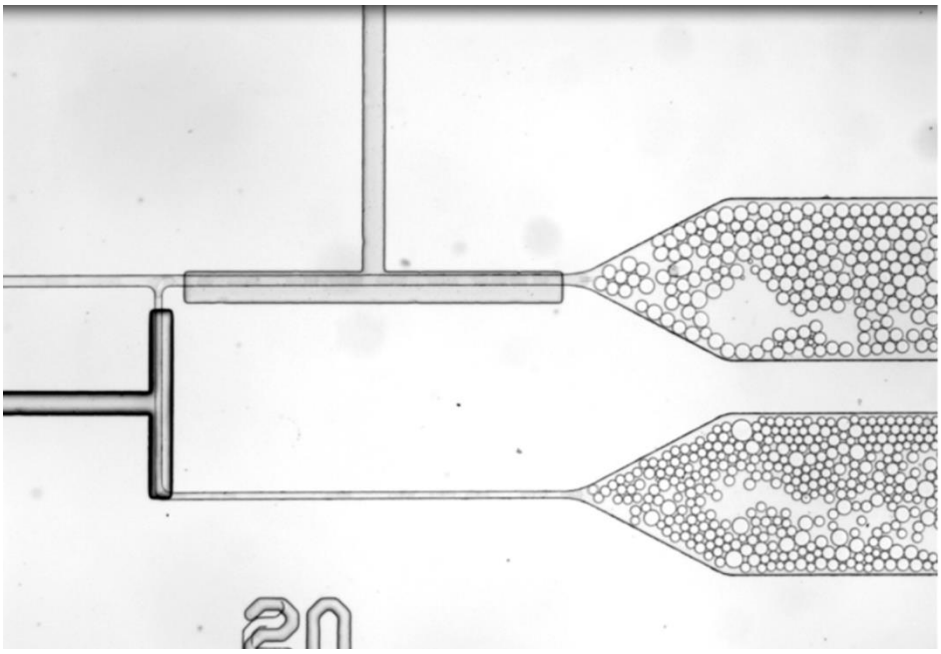


Figure 40: Multilayer microfluidic sorter with integrated micro valves.

After several attempts, we discovered that the overall dimensions of the microfluidic channels were not suitable for droplet-based microfluidics. Due to the shallow height of the flow channels, the fluidic resistance was increased by orders of magnitude, thus decreasing a reliable control of the fluid.

3.1.4.10 Single-layer device for droplet sorting

Single-layer membrane valves consist of a thin PDMS elastomeric membrane separating two micro channels. By pressurising the valve micro channel, the thin membrane deflects and constricts the neighbour channel and changes the hydrodynamic flow. Valve actuation relies on different parameters such as PDMS flexibility and membrane thickness. The fabrication of high-aspect ratio structures is required to produce a sufficiently flexible membrane capable of clogging the neighbour channel (Fig. 41).

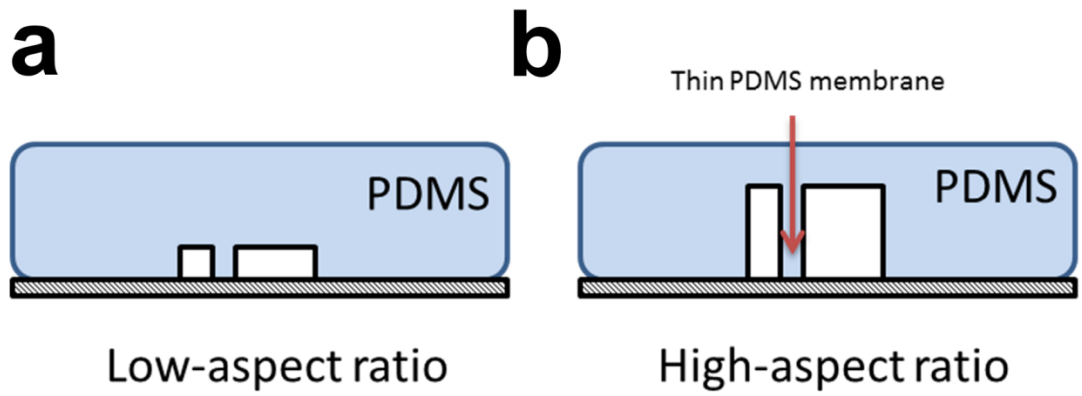


Figure 41: Microfluidic device with (a) Low-aspect ratio and (b) High –aspect ratio channels.

Previous work by Abate *et al.*,[211] demonstrated the fabrication of a microfluidic device capable of sorting droplets by elastomeric valve actuation. Unlike the other methods, the fabrication of this device can be performed in a single step by standard soft-lithographic techniques.

Therefore, we decided to reproduce the same device due to the ease of fabrication and high reproducibility. Preliminary prototypes were fabricated and tested (Fig. 42).

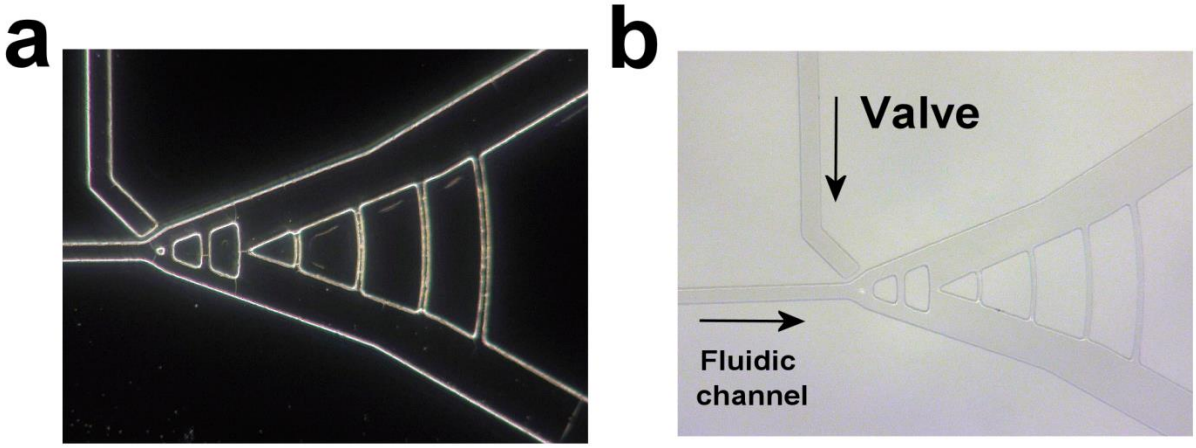


Figure 42: (a) SU-8 master of a single-layer valve sorter. (b) PDMS replica of a single-layer valve sorter.

Different photoresists and lithographic protocols were tested to optimise the resolution of the final thin membrane. However, the fabrication of a high-aspect ratio 20µm thin membrane is challenging and very precise lithographic tools are required. A few devices were successfully fabricated and tested, but other techniques are being developed to simplify the overall fabrication process.

The mechanism to sort droplets within a single-layer microfluidic sorter is based on the control of the hydrodynamic resistance of the flow. A population of droplets is injected into one of the inlets, whilst fluorous oil containing surfactant is injected into a second inlet. Oil spacing is required for sorting droplets individually. Then, in order to sort each individual droplet, an optical set up is required to detect the fluorescence/colour of each droplet, and transfer the information to an image recognition software that then triggers the actuation of the valve. In this case, the Y-junction was asymmetrically designed to drive the droplets towards a preferential channel (upper channel). When the valve is actuated, the thin membrane deflects to constrict the preferential channel and drive the droplets towards the other channel. Individual droplets can be sorted using this mechanism at a rate of 250Hz.

We were able to successfully fabricate a microfluidic sorting device with single-layer membranes. When a population of droplets was re-injected into the device, we were able to actuate the valve and selectively drive droplets towards the selected channels (Fig. 43).

However, we were not able to efficiently test the sorting capabilities of the device since we did have neither the right optical set-up, nor the adequate image recognition software to actuate the valve on demand. We are currently addressing some of these technical issues, which will be described in the next section.

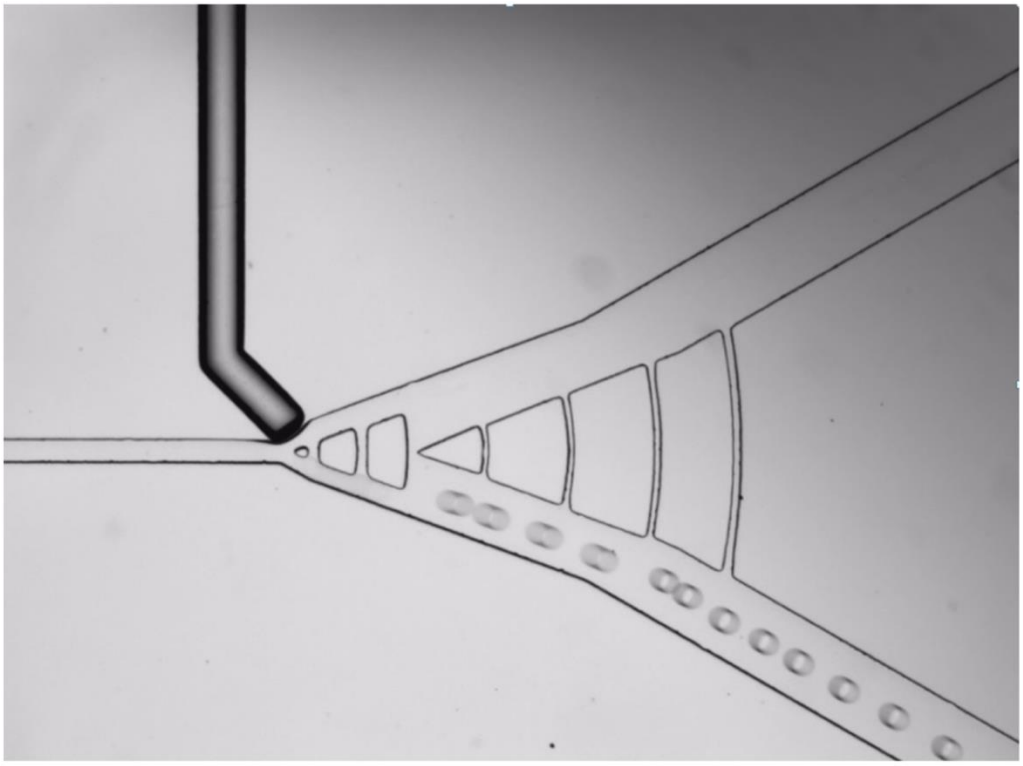


Figure 43: Single-layer microfluidic droplet sorter.

3.1.5 Ensemble sorting

As described in the previous section, we were able to fabricate a microfluidic droplet sorter based on single-droplet sorting. However, we started the development of a novel method of sorting: The ensemble sorter.

The ensemble sorter exploits the idea of sorting 'packets' or collections of objects rather than sorting object by object. This allows enrichment to occur over a number of cycles. In principle the ensemble sorter device has a 'sorting packet window', in which a packet of objects is flowed, in this case coloured droplets *(red and yellow).* The packet window evaluates the total amount of colour. If *red* droplets > *yellow* droplets, then the packet of droplets is flowed to one droplet reservoir, whilst if *yellow* droplets> *red* droplets, the packet of droplets is flowed to the opposite droplet reservoir.

Initially, since large packets of droplets are sorted at a time, the enrichment of one type of droplets into one reservoir would be small. However, as the droplet packets are re-sorted over iterative cycles, the overall enrichment will increase (Fig. 44). Traditional techniques for single-droplet sorting (FADS) are capable of sorting individual droplets at 1 KHz.[191] This requires very precise control of the droplet spacing and relies on electrophoretic actuation. By using our approach, we aim to be able to enrich heterogeneous droplet populations within iterative sorting cycles.

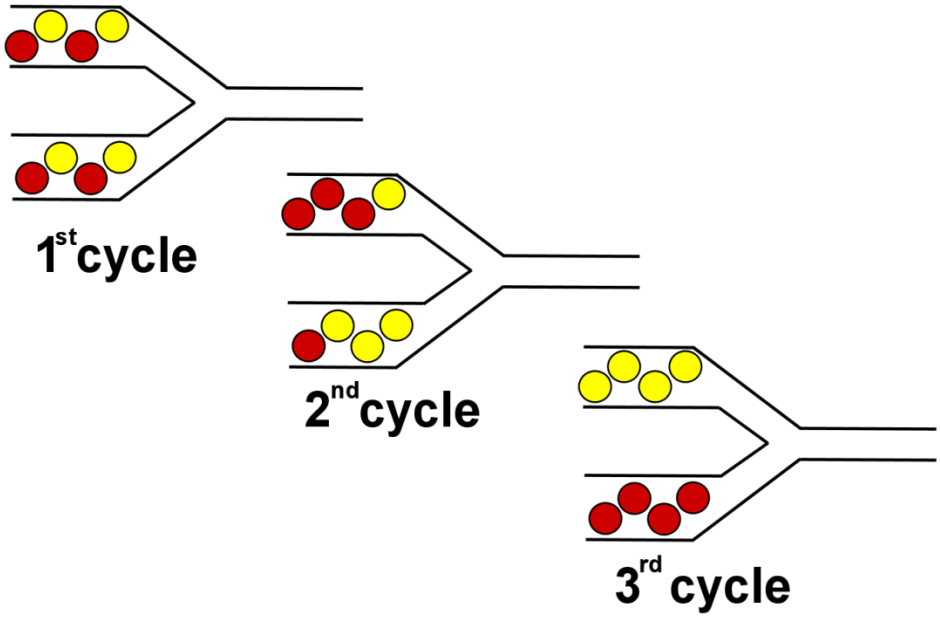


Figure 44: Object distribution based upon iterative sorting cycles. Over cycles, objects will be separated into different reservoirs.

A novel device was designed and fabricated to perform a series of sorting steps to select and isolate different subpopulations of droplets. The device symmetry allowed us to repeat sorting cycles by flowing subpopulations of droplets back and forth (Fig. 45).

Several designs were tested to improve the valve actuation. As mentioned in the previous section, fabricating 20µm PDMS membranes is difficult and hard to achieve. For this reason, we decided to add a second valve for each channel to constrict it from both sides. 30µm PDMS membranes were fabricated and valve actuation was successful.

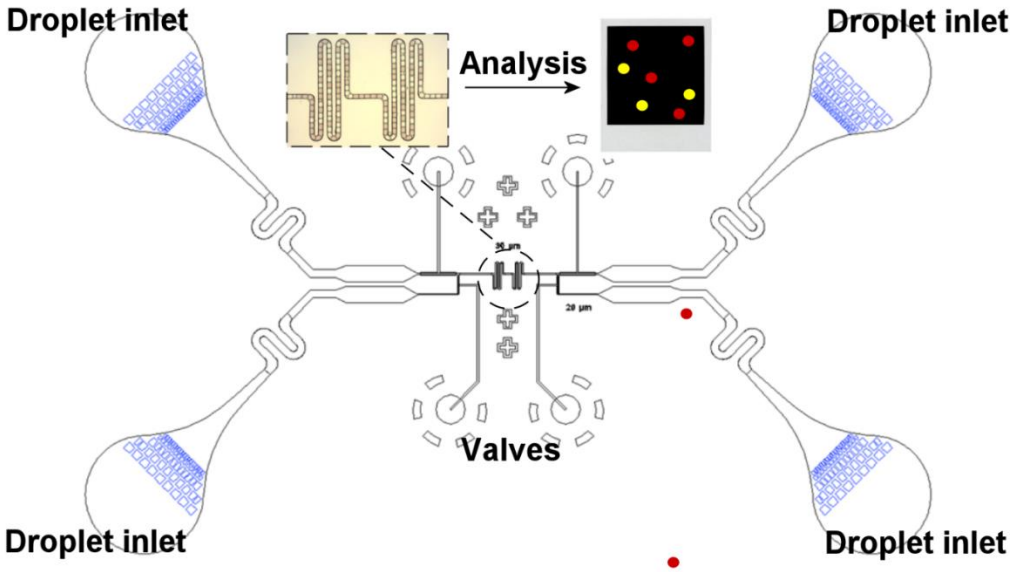


Figure 45: Schematic diagram showing the overall design of an ensemble sorter.

Initially, two different subpopulations of droplets are generated and re-injected into one of the inlets of the sorting device.

Figure 46: Red and yellow microdroplets flowing through one reservoir.

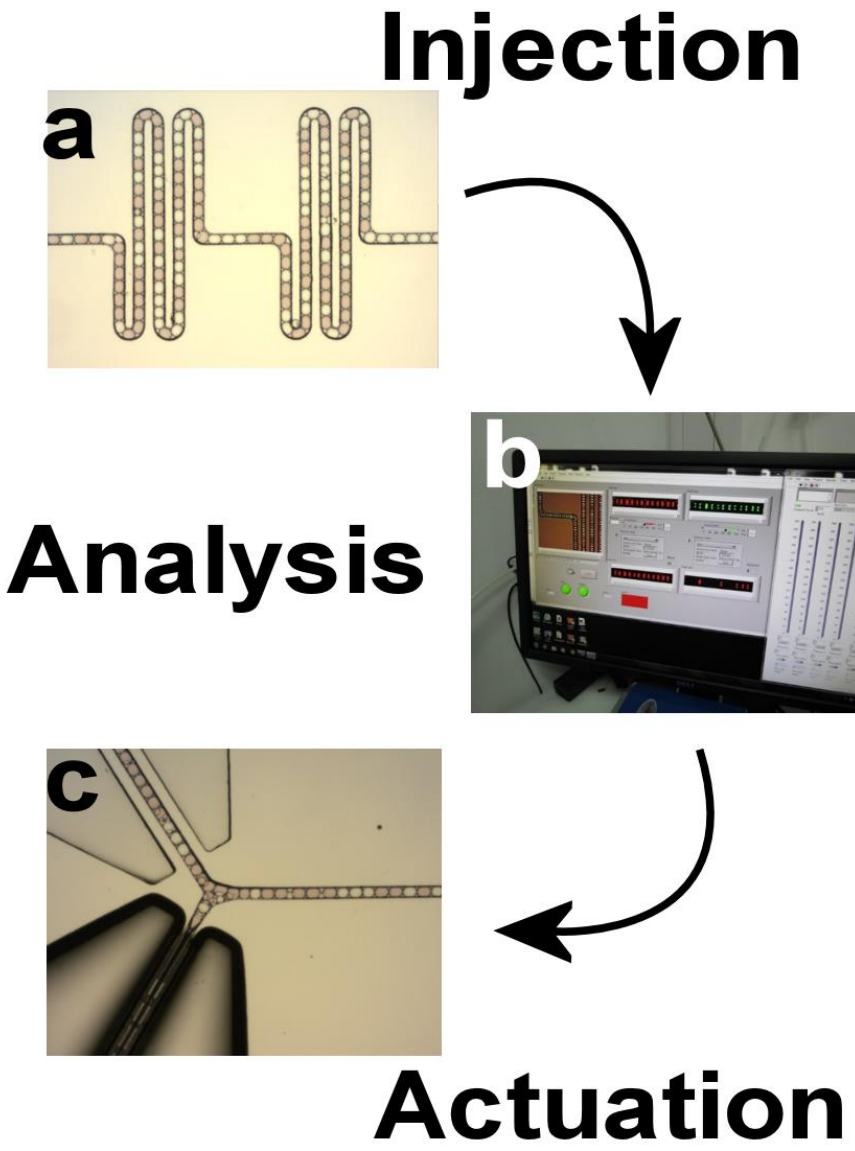


Figure 47: (a) Microdroplets flowing through the analysis point, (b) Image recognition software analysing the colour distribution for a determined number of microdroplets and actuating the valve as a function of the result and (c) Elastomeric Valve closed.

As illustrated in Fig.46, two different populations of droplets (red and yellow) are re-injected into one of the device inlets. Then, they are transferred toward the analysis window, a small channel serpentine where individual droplets of each colour are counted by an image recognition software (LabView). If the number of *red* droplets > *yellow* droplets, the software triggers the actuation of one of the valves, thus allowing the transfer of the droplet packet towards one reservoir. The opposite valve is triggered if the number *yellow* droplets > *red* droplets (Fig. 47).

So far, we have demonstrated the single-step fabrication of a robust microfluidic ensemble sorting device. Droplet re-injection, image recognition and valve sorting have been achieved. However, multiple automated selection cycles have to be performed to improve the sorting rate and compare the results with standard microfluidic droplet sorters. This is an on-going project that is now being finished with input from other researchers from the group.

3.1.5.1 Module integration

In the previous section, the design and fabrication of individual microfluidic modules was described. Droplet generation, droplet fission and droplet re-injection were successfully achieved, as well as the fabrication of micro-mechanical valves within double and single elastomeric layers.

However, efficient droplet fusion and sorting were not successfully achieved due to the limitations of our optical set up and the requirement of image recognition software. Nonetheless, a device containing all the different modules was designed and fabricated (Fig. 48).

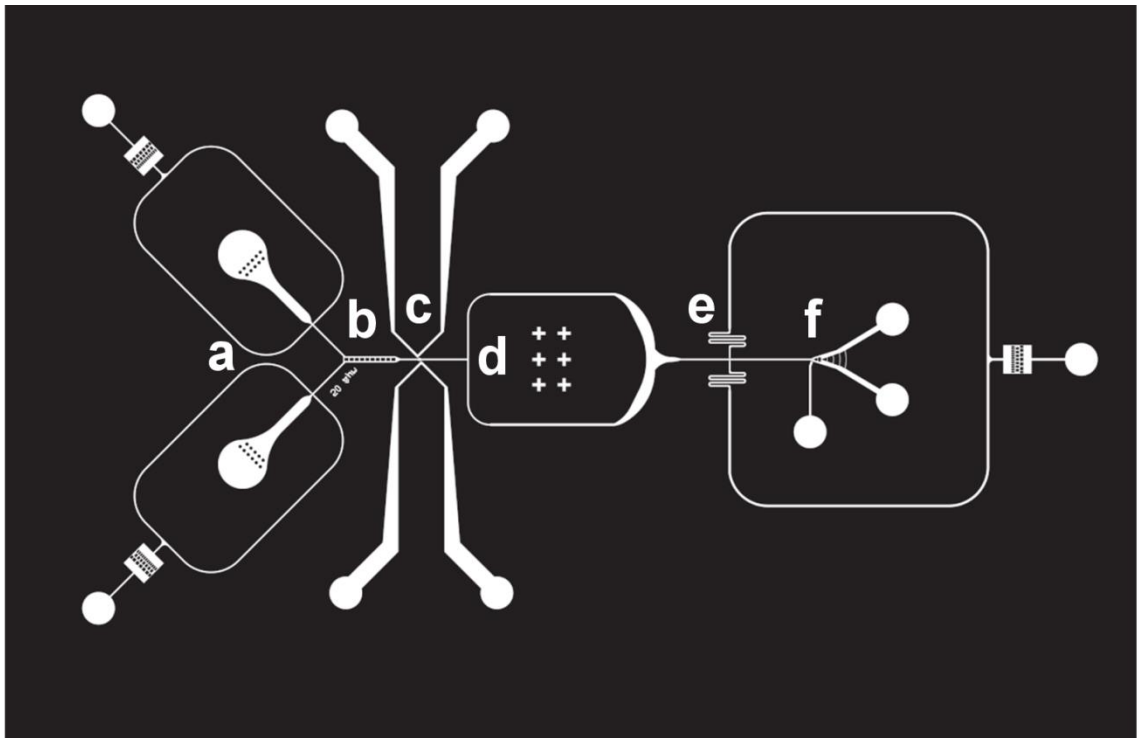


Figure 48: Photo mask of a microfluidic device containing: (a) Droplet generator/re-injector module, (b) Rail-road like channel for droplet synchronisation, (c) Droplet fuser module, (d) Droplet splitting module, (e) Droplet oil-spacer and (f) Single-layer valve sorting module.

With the device configuration represented in Fig.48, we could possibly perform all the unit operations described before once the fusion and the sorting step were optimised.

However, in order to repeat iterative cycles of pairing-fusion-splitting and sorting, we would need the selected droplets at the sorting step to be re-injected into one of the inlets at the beginning of a new cycle. Previous studies on enzyme evolution in droplets overcome this problem by collecting the selected droplets off-chip, and re-injecting them every cycle (this is because most of these reactions require incubation, which is complicated to integrate with the other modules on-chip). The initial aim of this project was to fabricate a microfluidic chip capable of performing all these unit operations autonomously, without external intervention, for a large number of cycles. A critical problem arises when droplets need to be transferred from the outlet to the starting point (without re-injecting them manually). Our first approach was to use valves to open and close channels to selectively drive the droplets to the desired reservoirs. As seen in the previous sections, we successfully achieved the fabrication and operation of such valves, and showed their performance by driving droplets towards different reservoirs in the Ensemble Sorter.

By using this approach, we could be able to transfer the selected droplets during the sorting step into the inlet reservoir. Nonetheless, a number of problems would arise from this operation. For correct droplet synchronisation, individual droplets need to be equally spaced when re-injected. This requires the droplets to be completely packed before spacing them with oil. If we were transferring droplets from the outlet to the inlet reservoir, these would have to be packed by removing the excess oil before the start of a new cycle. This step would require the addition of an oil ‘drain‘ feature in each one of the inlet reservoirs, and the use of an extra syringe pump that pulls the oil when the droplets are transferred.

Also, in order to ensure the correct performance of all the different unit operations, the synchronisation of droplets before the fusion steps would have to remain constant all the time. Small errors on the droplet pairing would lead to the absence of fusion and splitting of the droplets. To solve this problem, a feedback-control mechanism would be required to autonomously adjust the flow-rate of the inlets when droplet pairing is unsynchronised. This could be accomplished by analysing the size of the droplets after the fusion step: if the droplets keep their original size after the fusion step, it means that the pairing was not successful, and therefore needs to be adjusted until the size of the droplets is twice the size compared to the initial ones.

In addition to the previous mechanism, another analysis point would have to be placed before the sorting junction. This means that two different optical analyses would have to be performed simultaneously within the same chip, which constitutes a problem due to the field of view limitations of the camera.

3.1.6 Section summary

During the first stages of this project, the microfluidic lab facilities within the Cronin Group were settled. One year was required to build the cleanroom, install all the equipment and optimise the protocols for the fabrication of microfluidic devices. Once that was accomplished, we were able to focus on the design, fabrication and testing of the different modules for a microfluidic chip capable of assisting chemical evolution. As stated in the previous section, each one of the individual modules required were fabricated and tested, as well as a single microfluidic chip integrating all the different modules. Whilst, droplet generation, re-injection, splitting and valve integration were successfully achieved, droplet fusion and sorting were not completely accomplished due to the technical limitations of our set up.

At the latest stages of this project, when the technical difficulties were preventing further progress towards the initial objective, we decided to approach the formation of minimal living systems from a different point of view. Whilst we were trying to achieve the development of a microfluidic platform to assist chemical evolution, we had not specifically selected the chemical inputs for our experiments.

In an ideal scenario, we would start from a library of encapsulated reactive monomers that, upon mixing, would react to produce polymers with catalytic or replicating properties. Then, the droplets containing those polymers with improved properties would be selected by a sorting mechanism. Finally, the selected droplets would be driven towards the inlet to start a new cycle. Over iterative cycles, we would observe an increase in the complexity of the molecules and their catalytic or replicating properties.

Whilst directed evolution of enzymes has been previously reported using microfluidics, the starting materials used in these systems are highly evolved proteins with very specific catalytic properties.[189,210] Likewise, the directed evolution of RNA enzymes has also been demonstrated in microfluidics.[212,213]

The first step in our set-up involved the formation of functional polymers upon mixing compartmentalised reactive monomers. However, the synthesis of such polymers from their monomeric subunits constitutes one of the greatest challenges in Prebiotic Chemistry. Also, attempting to perform these reactions within microdroplets would limit their performance, since processes such as strong heating or evaporation are not compatible with these types of devices. However, the use of microfluidic technology developed during this project has potential in studying the role of compartmentalisation in prebiotic experiments. For example, the encapsulation of reagents within microdroplets has shown that thermodynamically unfavourable reactions can be enhanced by compartmentalisation at the micro scale due to reactant adsorption-diffusion mechanisms.[214] Also, it has been shown that microfluidics can be employed to generate and manipulate compartments which can serve as models of prebiotic enclosures.[215]

Given that the formation of biopolymers that could be later selected by directed evolution was the major impediment of this project, the focus was then placed on understanding the limitations of these processes. At the time this project was being developed, a Miller-Urey type experiment was being run in our laboratory. This inevitably led to consider the formation of peptides from the amino acids produced in this spark-discharge experiment and also inspired me to pursue and understand what have been the limitations in synthesising the building blocks of life, and their further polymerisation into relevant functional macromolecules.

3.2 Abiotic Peptide Synthesis

During the early stages of this project, we performed an extensive literature review on amino acid oligomerisation under abiotic conditions. Whilst a large number of scientific articles have been published in the last 50 years, studies have explored unactivated peptide synthesis on clays[216], minerals[124], at air–water interfaces[96], on metal oxide surfaces[217] and under hydrothermal conditions[101,104], resulting in very low yields (typically <1%) of oligomeric products (where n>3). Alternative approaches have shown the synthesis of longer peptides through the activation of amino acids.[111]

Nonetheless, the use of activating reagents is controversial in this type of syntheses due to their doubtful prebiotic significance. Therefore, we decided to explore the formation of peptide bonds from unactivated amino acids. The aim of this project was to achieve the formation of polypeptides with functional properties from amino acids produced in Miller-Urey type experiments, thus showing the transition from very simple precursors to complex molecules with functional properties.

Inspired by a systems chemistry approach, we decided to build an automated platform capable of exploring the effect of various parameters simultaneously under iterative cycles: The APS (Abiotic Peptide Synthesiser).

This chapter summarises the different APS versions that were explored to study the formation of peptide bonds under simple programmable conditions.

3.2.1 The initial APS platform

3.2.1.1 APS flow system

A recent study[214] has shown that thermodynamically unfavourable reactions can be enhanced by compartmentalisation at the micro scale due to reactant adsorption-diffusion mechanisms at the droplet interface. This study also suggests that compartmentalisation could have promoted unfavourable prebiotic reactions. Following this, we wanted to explore the effect of compartmentalisation on the formation of peptides from unactivated amino acids. The first version of our automated system was designed to generate aqueous droplets containing different mixtures of amino acids and minerals that could be concentrated and re-cycled in an automated fashion. Our platform was built by connecting a series of programmable syringe pumps that allowed in-cycle condensation reactions and re-cycling of the final products.

However, fabricating fluidic devices where aqueous droplets can be selectively evaporated in a controlled fashion is challenging. Whilst previous work has demonstrated the use of microfluidics to concentrate aqueous solutions to study phase diagrams of different solutes[218], the controlled evaporation of droplets within a fluidic system is hardly possible.

Therefore, we finally decided to encapsulate mixtures of amino acids and minerals within droplets to ensure proper mixing, followed by the evaporation of the bulk reaction mixture in a separate flask to promote the formation of oligomeric products.

By repeating the same unit operations over a determined number of cycles, we expected the formation of oligomeric products to vary as a function of the number of cycles and the chemical inputs (Fig 49).

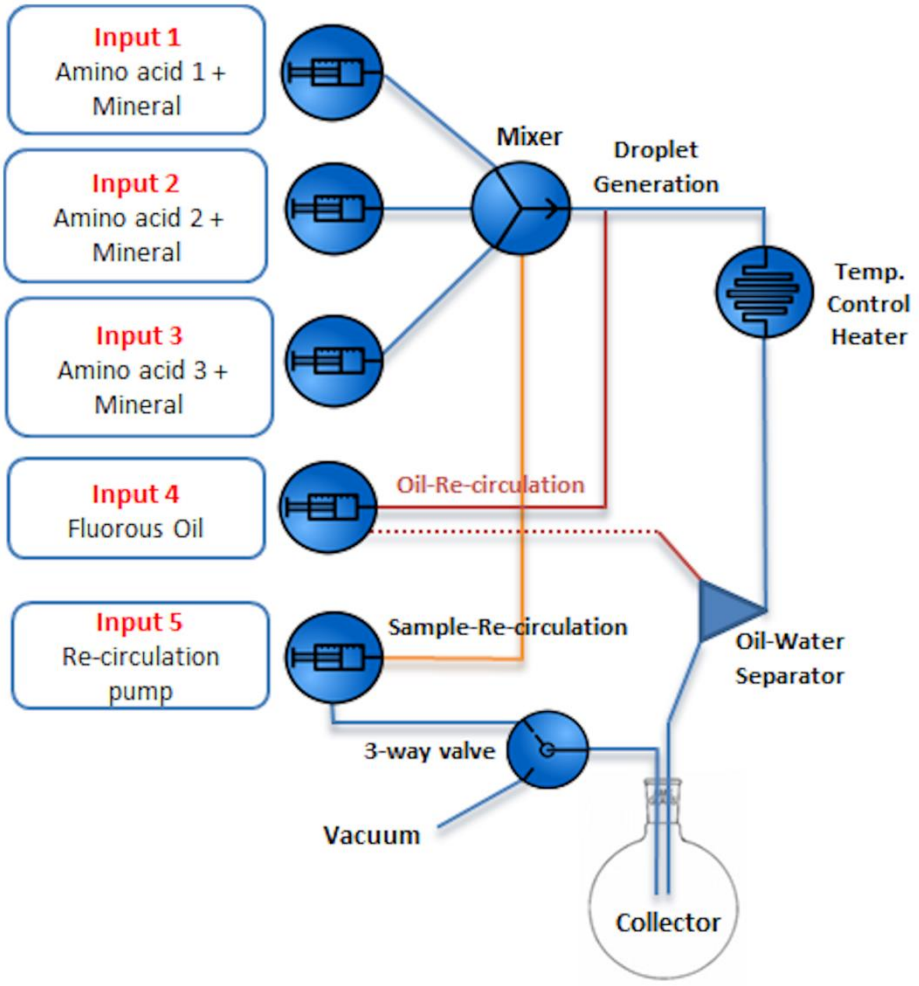


Figure 49: Schematic diagram showing the chemical inputs and unit operations performed in the system.

Step 1: A set volume is pumped from the inputs 1, 2 and 3 into the mixer. At the same time, the same volume is pumped from input 4.

Step 2: Segmented flow is generated within a T-junction mixer and goes through a heated coil at 90 °C.

Step 3: The aqueous phase is separated from the oil phase when the segmented flow goes through a separator device. The aqueous phase is collected into a separate container.

Step 4: The collected fraction is heated at 90 °C under vacuum for the condensation reaction.

Step 5: The concentrated fraction is recycled into the mixer and a new cycle starts

3.2.1.2 Fabrication of a water-oil separator device

Initially, we had to fabricate a device that could separate water-oil emulsions. By taking advantage of the 3D-printing technology being developed in our research group, we designed and fabricated a 3D-printed device containing a hydrophobic PTFE membrane capable of performing continuous-flow separation of oil-water mixtures.

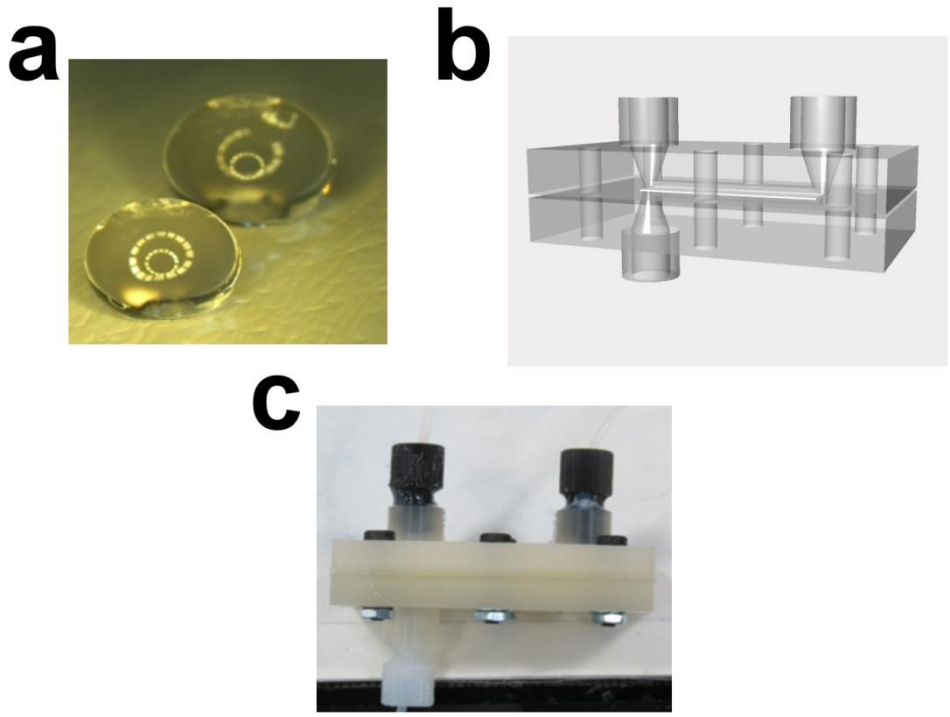


Figure 50: (a) Two water droplets on a 0.45µm pore PTFE membrane. (b) Rendered structure of a 3D model of the separator device. (c) Final 3D-printed separator device.

The device consists of two separate parts that are bonded together by dissolving their surface with dichloromethane (Fig. 50). The hydrophobic PTFE membrane is aligned between the two parts before bonding. The fabrication of a device can be rapidly achieved within a few hours (from the initial design until the final prototype). Several designs and printing parameters were tested to improve the separation performance and avoid leakage problems. Despite the ease to fabricate a device, due to the porosity of the material and the non-optimised bonding strategy, leakage was frequent at moderate pressures.

Alternative materials with less porosity and higher chemical resistance, such as polypropylene, were also tested.

We were able to fabricate a separator device by using soft lithography. In this case, we used PDMS (a silicon-based polymer), which is the most common material for the fabrication of microfluidic devices (Fig. 51). The channel pattern was 3D-printed and used as a master for PDMS replica molding. Standard soft lithographic techniques were used to cast, cross-link and peel-off the PDMS slabs from the masters. In this case, different PDMS mixing ratios were used to ensure a solid bonding between the two layers.

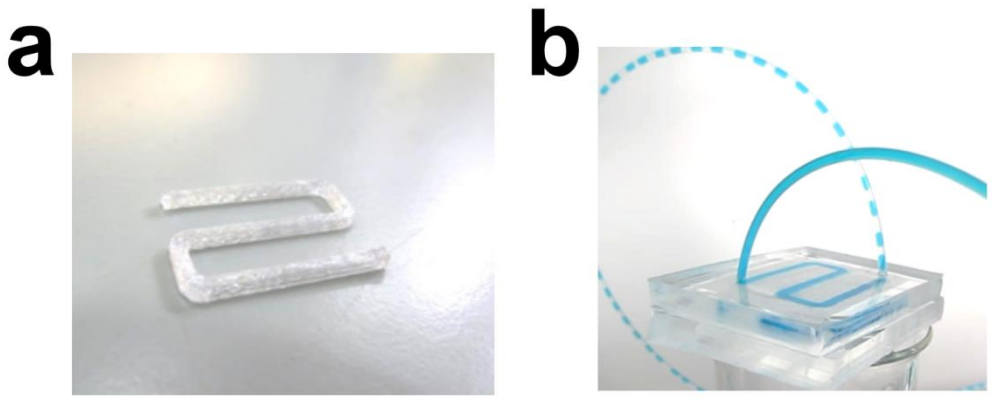


Figure 51: (a) 3D-printed channels (b) Final elastomeric PDMS separator device.

3.2.1.3 Preliminary experiments

Once all the individual modules of the system were tested and functioning, we performed a series of preliminary experiments to ensure the correct performance and stability of the system. Glycine, alanine and lysine were the amino acids chosen, whilst potassium Montmorillonite was chosen as a mineral catalyst. In 1978, Lahav *et al.*[216] demonstrated the formation of glycine oligomers by repeatedly exposing a glycine solution to hydration-dehydration cycles in the presence of a clay catalyst. This was the first example where polypeptides were being formed over successive drying and rehydration of unactivated amino acids. However, very low yields were obtained (>1% where n≥3). In light of these results, we wanted to use our automated approach to investigate a broader range of conditions (number of cycles, dehydration time, and temperature).

Initially, all our experiments contained a constant amount of K-Montmorillonite, a clay mineral that has been extensively studied and is known to promote the polymerisation of amino acids into peptides.[219]

Montmorillonite is a smectite group mineral with chemical composition $(Al,Mg)_2(Si_4O_{10})(OH)_2 \cdot nH_2O$. It has a high surface area and a high exchangeable cation capacity. It has been proposed that Montmorillonite catalyses the formation of peptide bonds due to the absorption of amino acid monomers via hydrogen bonding between the positively charged amino groups and negatively charged surface sites (i.e., Lewis base sites) on the montmorillonite surface, in which the Lewis base site acts as a catalyst for peptide formation.[220]

In order to extract the possible oligopeptides formed during the condensation reaction, we diluted the reaction product with a 0.1M $CaCl_2$ solution.

The objective of this preliminary test was to study whether increasing the number of dehydration cycles had an effect on the amino acid condensation reaction.

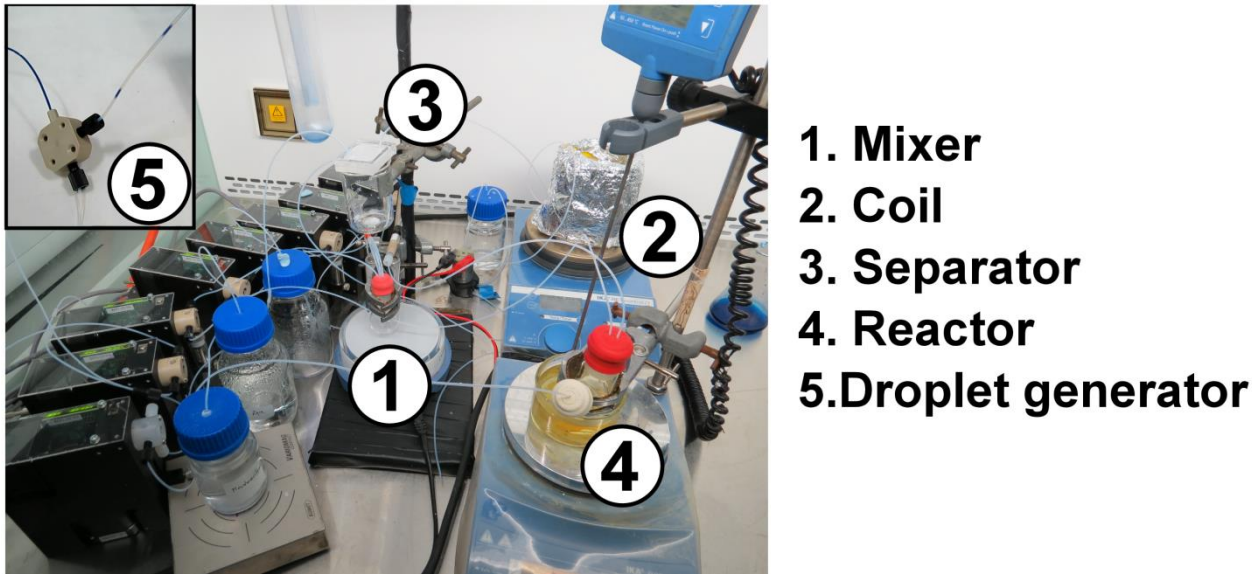


Figure 52: Picture showing the APS flow set-up.

The unit operations of each cycle were controlled by LabView. We designed a program capable of controlling the flow rate of addition, volume and cycle time of each individual experiment within a continuous loop. Each one of these values could be easily modified between experiments, as well as the total number of cycles. Before starting the experiment, the desired values were entered in the program and the experiments were left running autonomously over a specific period of time (depending on the number of cycles).

Low concentrations of starting materials were chosen to control the successive feeding of monomers into the reactor between each dehydration cycle. The starting parameters for the first set of experiments remained constant, except the number of cycles. Each experiment was running autonomously for the specified number of cycles. The reaction products were then collected and stored in the -20 °C freezer before analysis.

All the reaction products were analysed by Size Exclusion Chromatography.

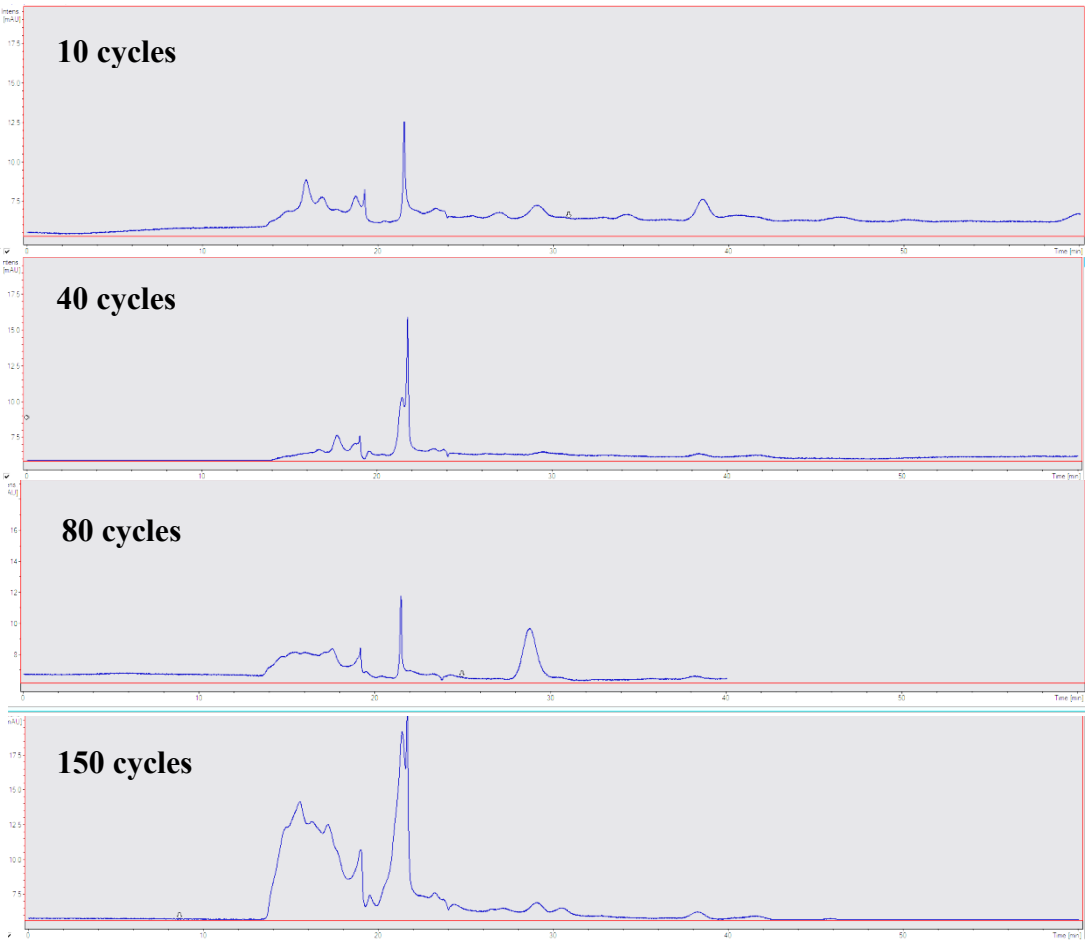


Figure 53: HPLC chromatogram recorded at 254 nm for the first set of experiments.

In size exclusion chromatography, smaller molecules elute later than larger ones. Based on this principle, we observed how successively increasing the number of dehydrating cycles led to the formation of larger species in our system (Fig. 53).

However, further analysis was required to find out the oligomer length of the larger species identified by this separation. Therefore, the sample that ran for 150 cycles was analysed by mass spectrometry. Relevant information could not be extracted from the spectrum due to the large presence of peaks between 300m/z-800m/z that could not be assigned to a specific peptide sequence.

In order to simplify the analysis of our samples and demonstrate the formation of peptide oligomers, we then decided to perform the same reaction starting from a mixture of glycine and its dimer. If oligomeric products were obtained, they could be easily identified by mass spectrometry.

The APS system was set up to perform 200 hydration-dehydration cycles starting from glycine, glycylglycine and potassium Montmorillonite.

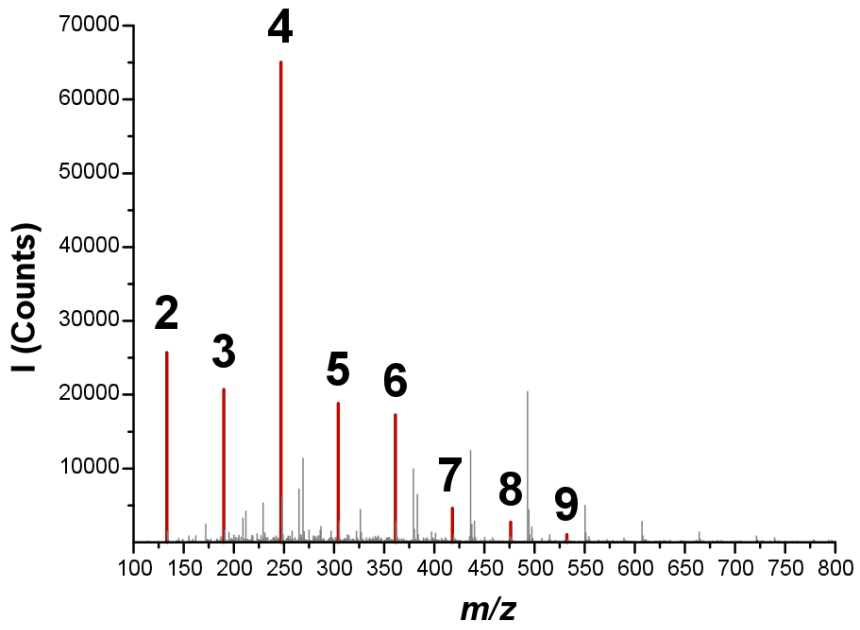


Figure 54: MS spectra of the solutions resulting from washing the product of the APS reaction with 0.1M $CaCl_2$ after 200 cycles.

Relevant peaks could be then assigned for glycine oligomers between 1 and 9 units (Fig. 54). We then performed tandem mass spectrometry (Fig. 55) to analyse the fragmentation pattern resulting from the peaks obtained in Fig. 54.

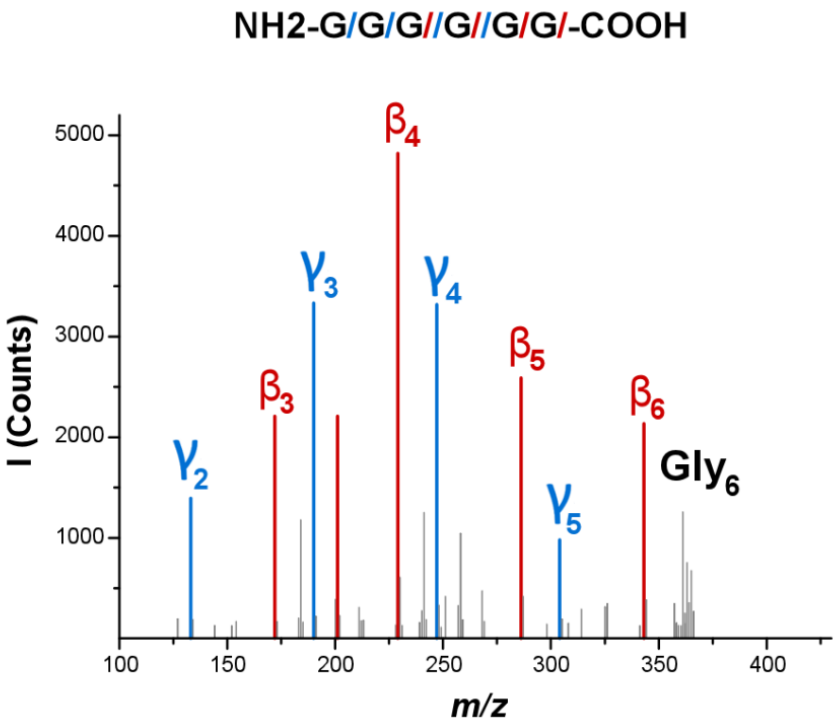


Figure 55: Typical SEC-MS/MS spectrum of one such species (Gly_6, *m/z* = 361.15); the peptide structure is confirmed by fragmentation to β- and γ-series product ions of masses predicted by theory for CID (Collision Induced Dissociation) of Gly_6.

By analysing the fragmentation pattern, we observed that the resulting peaks corresponded to γ and β oligoglycine fragmentation peaks. In these preliminary experiments, we demonstrated the formation of glycine oligomers up to 9-mers when analysed by MS.

However, having an integrated droplet generator that required the continuous pumping of mineral oil through our fluidic system was causing continuous technical problems. Due to occasional failures in the water-oil separation module, oil fractions were often transferred to our reaction flask where the dehydration reaction was taking place. Therefore, we decided to remove the droplet generator module from our original set up. Later experiments without the droplet generator module produced similar results, which confirmed that in this experiment the encapsulation of reagents in droplets did not contribute to the improvement of the oligomerisation reaction.

The initial set up was designed to perform a single reaction at a time. This was accomplished by pumping an amino acid solution into a heated round bottom flask under vacuum to speed up the condensation reaction.

Since the configuration of the set up was limiting the number of individual experiments that could be performed in parallel, we decided to change the way the samples were dehydrated.

Aerosols have been suggested as prebiotic chemical reactors for the condensation of amino acids[97], and they represent a fast method for evaporating aqueous solutions when heated on a hot surface. Therefore, we decided to incorporate aerosol generators in our experimental set up. In this way, we fabricated a series of aerosol generators by inserting a dispensing needle into a barbed T-connector. Every single syringe pump in our set-up was connected to an aerosol generator through PTFE tubing, so 8 different experiments could be run in parallel.

As mentioned in previous research, the activation energy to bond a glycine monomer to a glycylglycine molecule is lower than the energy required to bond two glycine monomers.[76]

In our preliminary experiments, we were using a mixture of glycine and glycylglycine as starting materials to explore a feasible range of conditions by which glycine oligopeptides could be formed. However, our ultimate aim was to find an optimum set of conditions in which glycine oligomers could be produced starting from just glycine monomer.

A series of experiments were performed where a mixture of Gly, GlyGly and K-Montmorillonite was successively sprayed on a hot vial for a different number of cycles and dehydration times. In this case, Ion-pairing HPLC (IP-HPLC) was used to determine the formation of oligoglycine species (Fig. 56). We followed the method described by Rode *et al.* [95]

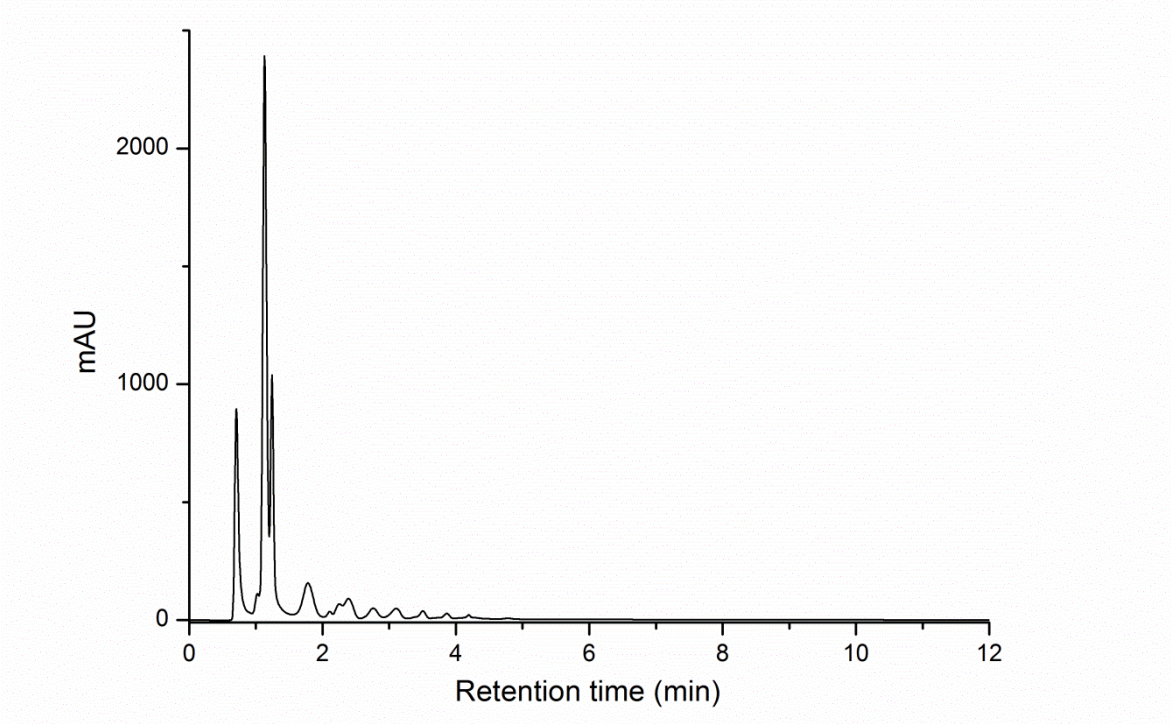


Figure 56: HPLC chromatogram recorded at 195nm for an experiment containing Gly, GlyGly and K-Montmorillonite after 500 dehydration-hydration cycles (8min dehydration time, 130C).

Whilst performing these experiments, a recent study on the effects of pH and temperature in dimerisation of glycine came to our attention. In this case, they found that under hydrothermal conditions the rate of glycine dimerisation is strongly dependent on the starting pH of the solution.[105] Glycine dimerisation proceeds more readily at high pH where unprotonated amine groups are more nucleophilic, and hence more readily attack partner carbonyls, whereas at neutral pH glycine monomers are zwitterionic and protonation of the amino lone pair prevents the reaction.

Therefore, we decided to run a set of experiments under the pH conditions described in this work. The first IP-HPLC results showed a dramatic increase in the overall area where the possible glycine oligomers were eluting.

At the same time, we were optimising an IP-HPLC method to analyse the glycine oligomers formed during this reaction. Glycine, diketopiperazine (DKP) and glycine oligomers (n=3-6) standards were successfully separated and quantified by this method. However, the first samples analysed by the reaction run under basic conditions (pH=9.8), produced a series of peaks that did not correspond to the standards. After obtaining the same poorly resolved chromatograms over a series of different experiments, we realised that the basic nature of the samples could be the cause for this problem. Initially, our samples were prepared for analysis by dissolving the dehydrated product with 8 ml of a 0.1% TFA aqueous solution.

The resulting pH of the solution after the addition of 8ml of a 0.1% TFA aqueous solution to the dehydrated sample was 9.2. Since the pH of the Ion-Pairing buffer used in the HPLC separation was 2.5, we decided to further decrease the pH of our sample before the analysis. In this way, the dehydrated samples were then diluted by adding 4 ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution. The final pH of the resulting solution was similar to the pH of the HPLC mobile phase.

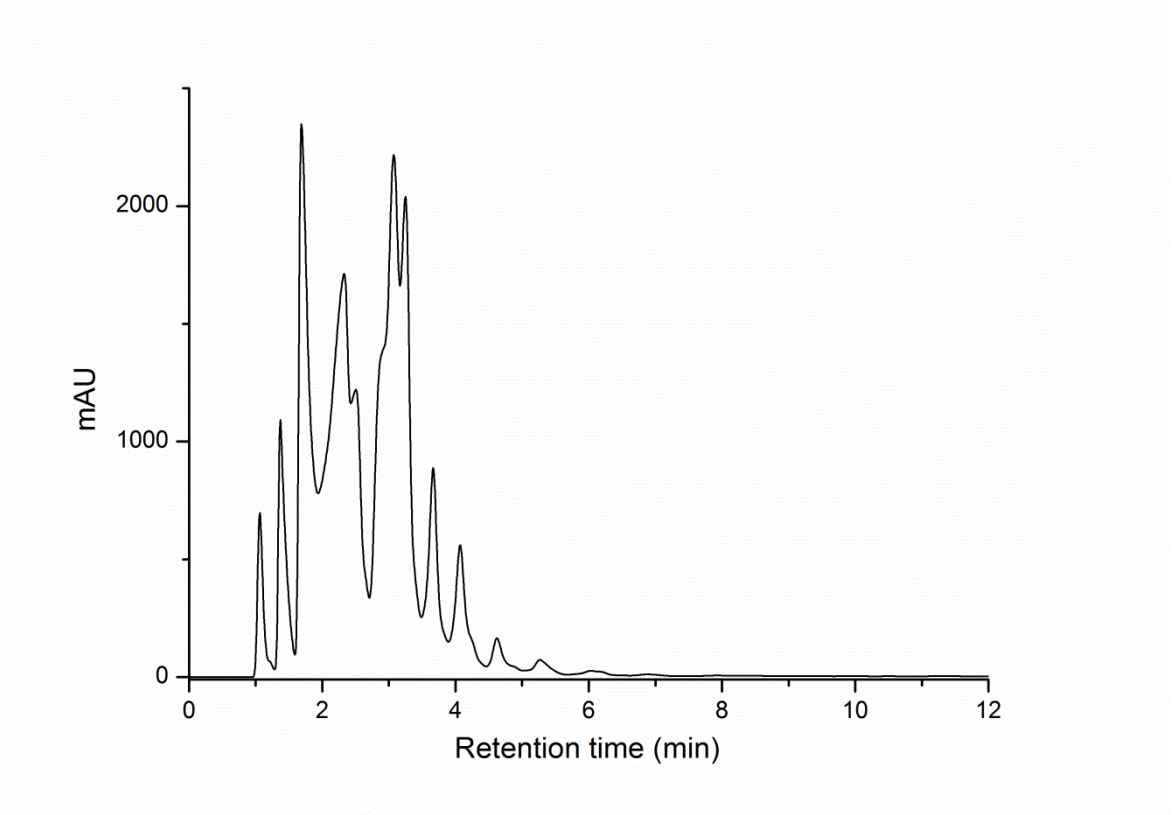


Figure 57: HPLC chromatogram recorded at 195nm for an experiment containing Gly, GlyGly and K-Montmorillonite after 50 dehydration-hydration cycles (5min dehydration time, 130C). In this case, the final product was diluted in 8 ml of a 0.1% TFA solution. Then, 100µl of this solution were added to 100µl of H_2O.

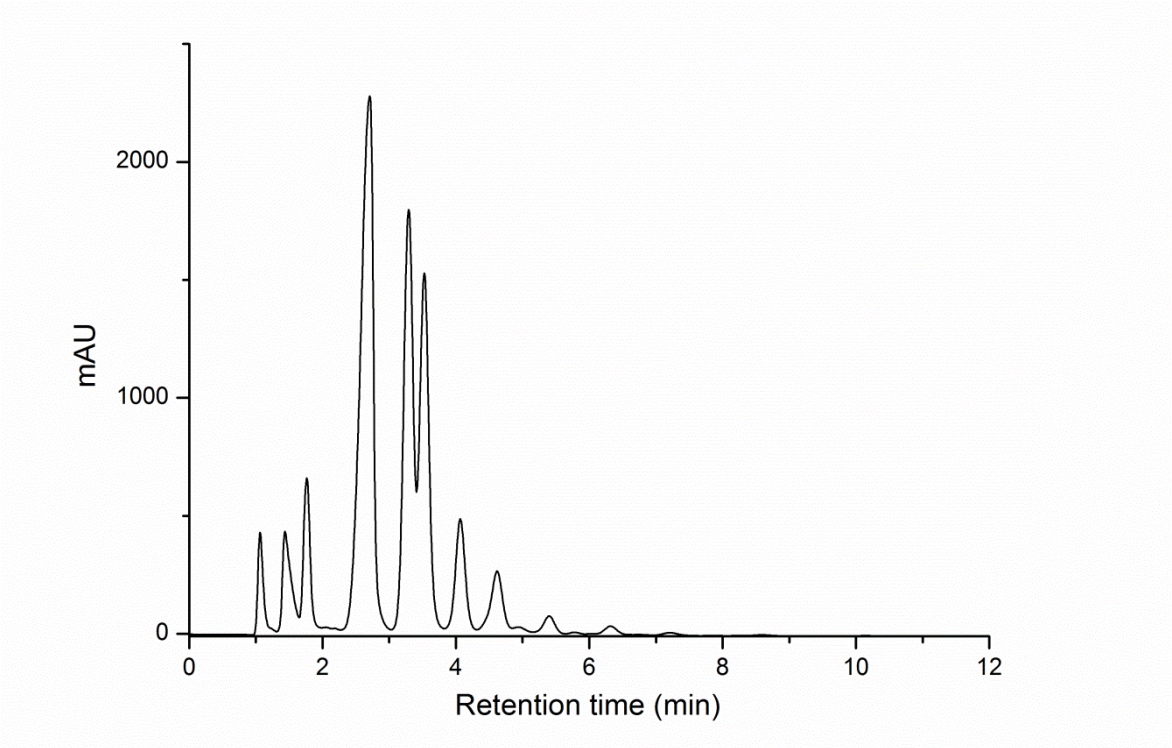


Figure 58: HPLC chromatogram recorded at 195nm for an experiment containing Gly, GlyGly and K-Montmorillonite after 50 dehydration-hydration cycles (5min dehydration time, 130C). In this case, the final product was diluted in 4ml of a 0.1% TFA solution. Then, 100µl of this solution were added to 100µl of 0.05M H_3PO_4.

As observed in Figs. 57&58, adjusting the final pH of the samples to the pH of the mobile phase dramatically increased the quality of separation in the chromatograms. Having established a standard protocol for the analysis of our glycine oligomeric products, we were now able to study quantitatively the formation of glycine oligopeptides under different experimental conditions.

In all the reactions run previously, the starting amino acid solution consisted of an equimolar mixture of Gly and GlyGly, along with Montmorillonite K-10. Under neutral pH conditions, we could not observe the formation of glycine oligomers when just glycine monomer along with MM K-10 was used as a starting material. Nevertheless, when the same starting solution was adjusted to pH=9.8, the formation of glycine oligomers up to hexamers could be easily observed by IP-HPLC.

In the initial stages of this set of experiments, we thought that the rapid evaporation of the amino acid solution via the formation of an aerosol was an essential feature to promote the formation of peptide bonds. However, when one of the experiments was left under dry heat overnight due to an error with the syringe pump, we observed a much higher conversion to glycine oligopeptides than in any experiment run before.

These results suggested that not only the pH but also extended dehydration times strongly promoted the oligomerisation reaction. In order to understand the effect of the dehydration time (between a large number of successive short dehydration-hydration cycles or a single long dehydration cycle), we decided to perform a standard reaction in the absence of an aerosol generator. In this reaction, the preparation of a dilute solution of starting material was performed by taking 350 μl from a 1M solution of glycine, to which we subsequently added 1 ml of a 1M NaCl solution and 2.55ml of (HPLC) water, and finally the pH was adjusted to 9.8 by adding 100 μl of 1M NaOH. The final solution was added to a 7ml glass vial and placed in a convective-flow oven at 130˚C overnight.

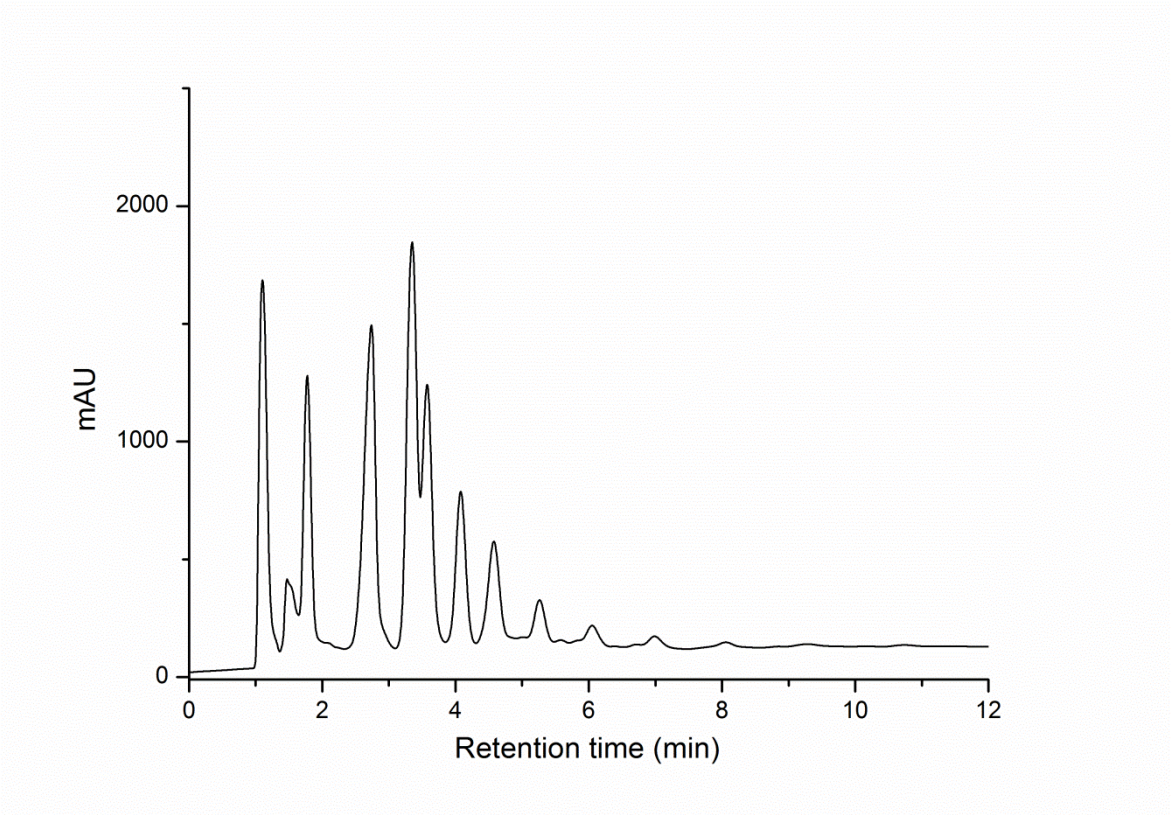


Figure 59: HPLC chromatogram recorded at 195nm for an experiment containing Gly after 1 dehydration-hydration cycle (15h dehydration time, 130C). In this case, the final product was diluted in 4ml of a 0.1% TFA solution. Then, 100µl of this solution were added to 100µl of 0.05M H_3PO_4.

These results (Fig. 59) showed that glycine oligomerisation proceeds more easily than expected under the right set of conditions. We realised that neither the addition of a clay catalyst nor the constant spraying of the sample greatly contributed to the improvement of the reaction yield under these set of conditions.

Having established that the formation of glycine oligomers was possible under 'one-pot' conditions, we decided to re-design our automated platform in order to perform bulk reactions (Fig. 60).

Figure 60: Latest APS system. The syringe pumps were connected to the vials through 1/16'' PTFE tubing. Glass reaction vessels were placed in the corresponding Drysyn hotplate inserts. Custom-made three-dimensional printed polypropylene lids with integrated holes were placed on each vial to connect them to the pumps, which facilitated the evaporation during the drying step. Water solutions were connected to each individual pump to deliver a given volume in each re-hydrating step.

3.2.2 The abiotic peptide synthesiser system

In this work, we set out to explore the formation of peptide oligomers under the simplest possible 'one-pot' reaction conditions. To do this, aqueous solutions of amino acid monomers were added to a hot empty glass vial and the water was removed by continuous heating. For subsequent dehydration-hydration cycles, further amounts of water were added to the reaction vials, and were again allowed them to completely evaporate. By taking this approach we aimed to explore the formation of peptide bonds with minimal chemical inputs. As we wished to investigate many different combinations of conditions (temperature, concentration of starting materials, pH, number of dehydration-hydration cycles), we re-designed the 'Abiotic Peptide Synthesiser' (APS) system in order to run several reactions in parallel (see Figs. 60&61), and automatically vary both the input and process variables.

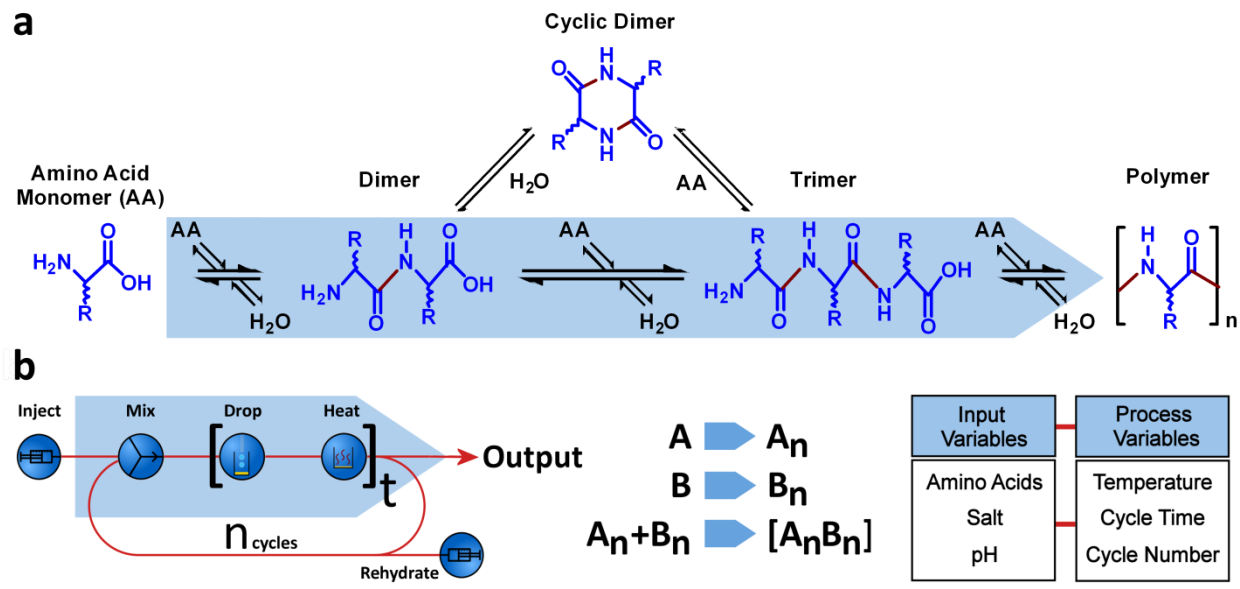


Figure 61: **The 'Abiotic Peptide Synthesiser' (APS).** (**a**) Schematic of peptide bond formation to give oligomers from amino acid monomers. (**b**) Process flow scheme for the 'Abiotic Peptide Synthesiser' (APS) system: amino acid starting material is added and dehydrated for a given time (t) then rehydrated/dehydrated for a given number of cycles (n), after which the product outputs are removed for analysis. Single amino acid monomers (A, B etc.) may be added or, alternatively, defined amino acid 2-3-mer fragments (A_n, B_n) can be used as starting materials.

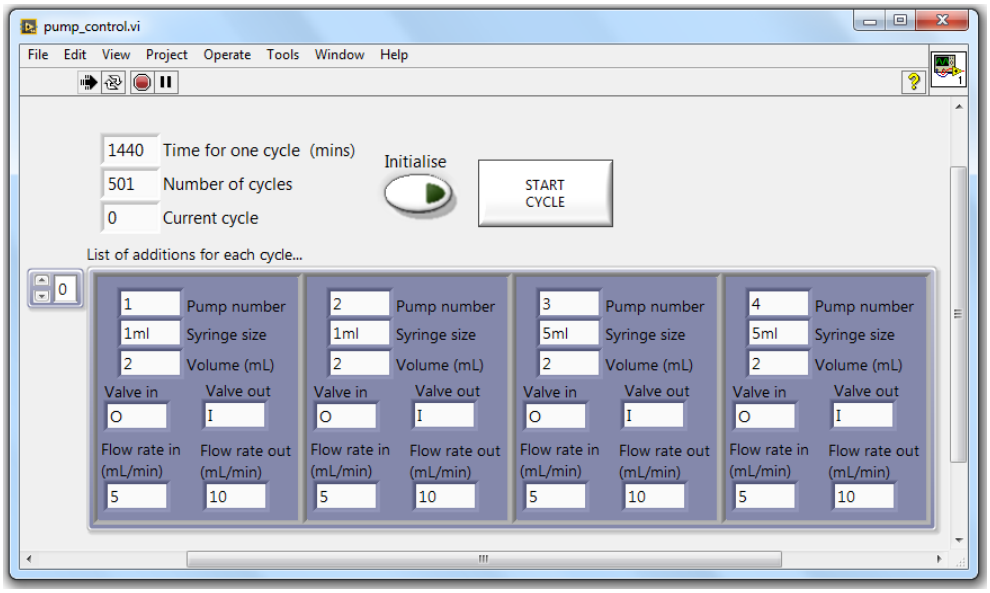


Figure 62: A screen capture image of the National Instruments LabVIEW control software user interface that controlled the array of individual experiments (see Figs Figure A1-A3for further details).

Reaction outcome was controlled by selection of input variables and process variables and the overall reaction time was simply the time per cycle × the number of cycles (Fig. 62).

A typical reaction run involved injection ($10ml \cdot min^{-1}$) of an aqueous solution of glycine (0.0875M) containing NaCl (0.25 M), pH adjusted with NaOH (to 9.8), into a pre-heated vial (T = 90 to 130 °C) which was then maintained at that temperature for approximately 15h, evaporating the solution to dryness (the 'dehydration step'). Subsequent cycles began with rehydration of the sample with 4ml of water (the 'hydration step'), and proceeded similarly; after the final cycle, the vial was allowed to cool to room temperature. To prepare a solution for analysis, 8ml of water was then added; typically some insoluble precipitate was also observed, which was analysed separately. After only one dehydration-hydration cycle we already observed oligomers in solution (up to Gly_{12}) in up to around 50% yield by Ion-Pairing High Performance Liquid Chromatography (IP-HPLC, see Figs. A4&A5) along with some cyclic glycine dimer (diketopiperazine, DKP; not included in yields). Once we had an automated system running and preliminary results clearly showing the formation of glycine oligomers, we decided to further confirm the nature of these products by a series of different analytical techniques.

3.2.2.1 Analysis of the products

Size Exclusion Chromatography-Mass Spectrometry (SEC-MS) was used to further confirm the identity of the oligomeric products and the presence of peptide bonds, revealing a range of oligomers which yielded tandem mass spectrometry (MS/MS) fragmentation patterns consistent with a peptide structure (see Fig. 63 a and b, and Fig. A8-A12). This was further corroborated by Infra-Red (IR) analysis of precipitate (consistent with oligomer standards, see Fig. 63 c and d) and proton Nuclear Magnetic Resonance Spectroscopy (NMR) analysis (see Figs. A13-A16).

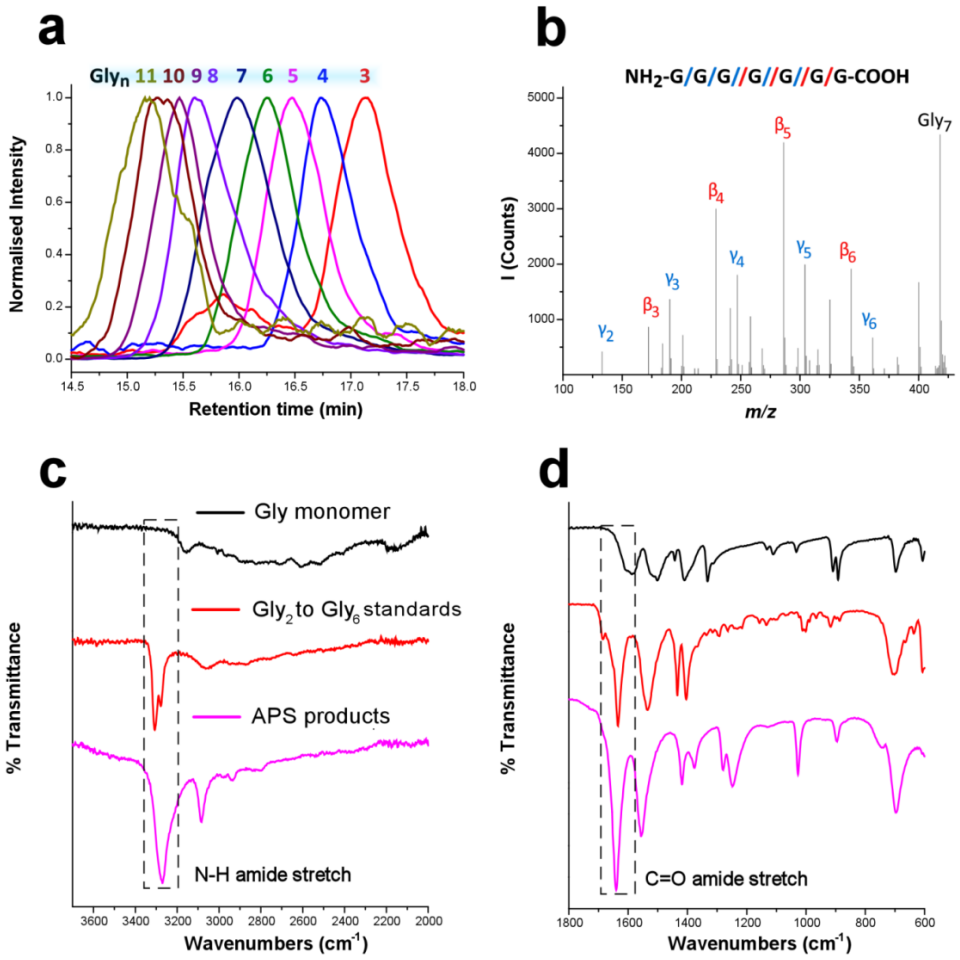


Figure 63: (**a)** Extracted Ion Chromatograms (EIC) for a series of glycine oligomers – from Gly_3 to Gly_{11} – produced in a typical monomeric glycine experiment on analysis by Size Exclusion Chromatography-Mass Spectrometry (SEC-MS), see Methods sections for details. (**b)** Typical SEC-MS/MS spectrum of one such species (Gly_7, *m/z* = 418.17); the peptide structure is confirmed by fragmentation to β- and γ-series product ions of masses predicted by theory for CID (Collision Induced Dissociation) of Gly_7 . Infra-Red showing the comparison of glycine monomer, oligomer standards, and the reaction products from the APS. c) IR in the region 3600 to 2000 wavenumbers showing the N-H amide stretch and d) IR in the region of 1800 to 600 wavenumbers showing the C=O amide stretch.

3.2.2.2 Exploring the parameter space

One important aspect of this work was the development of a robotic protocol to explore the parameter space for peptide bond formation. Our automated robotic system allowed us to automatically search the parameter space. To establish a feasible range of dehydration times for glycine oligomer formation, APS reactions were carried with dehydration times between 1 and 24 hours. IP-HPLC analyses of the resulting soluble products showed that longer reaction times led to progressively longer oligomer products; products up to Gly_{12} could be observed after a single dehydration-hydration cycle (Fig. 64). Furthermore, significant quantities of insoluble products were observed in some of these reactions. To produce a solution amenable to IP-HPLC analysis, these precipitates were washed with an aqueous solution of 0.1% *v/v* trifluoroacetic acid (TFA); while not dissolving all the material, it was possible to observe larger oligomeric glycine species in the fraction which was dissolved and that this comprised mostly of higher oligomers (n > 5), see Fig. 65b. While not a quantitative technique, Matrix Assisted Laser Desorption/Ionization (MALDI-MS) analysis of these insoluble fractions yielded evidence of oligomers > 20-mer (see Fig. 64).

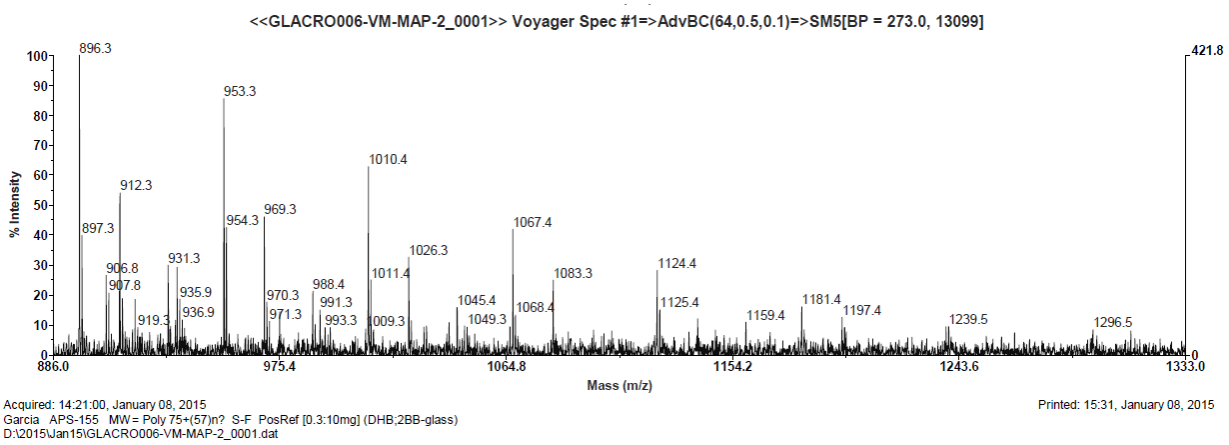


Figure 64: MALDI-TOF analysis of a solid fraction (zoom in high mass region). Note peak at 1181.4, corresponding to $[C_{40}H_{62}N_{20}O_{21}Na]^+$, the composition of the sodium ESI adduct of $(Gly)_{20}$.

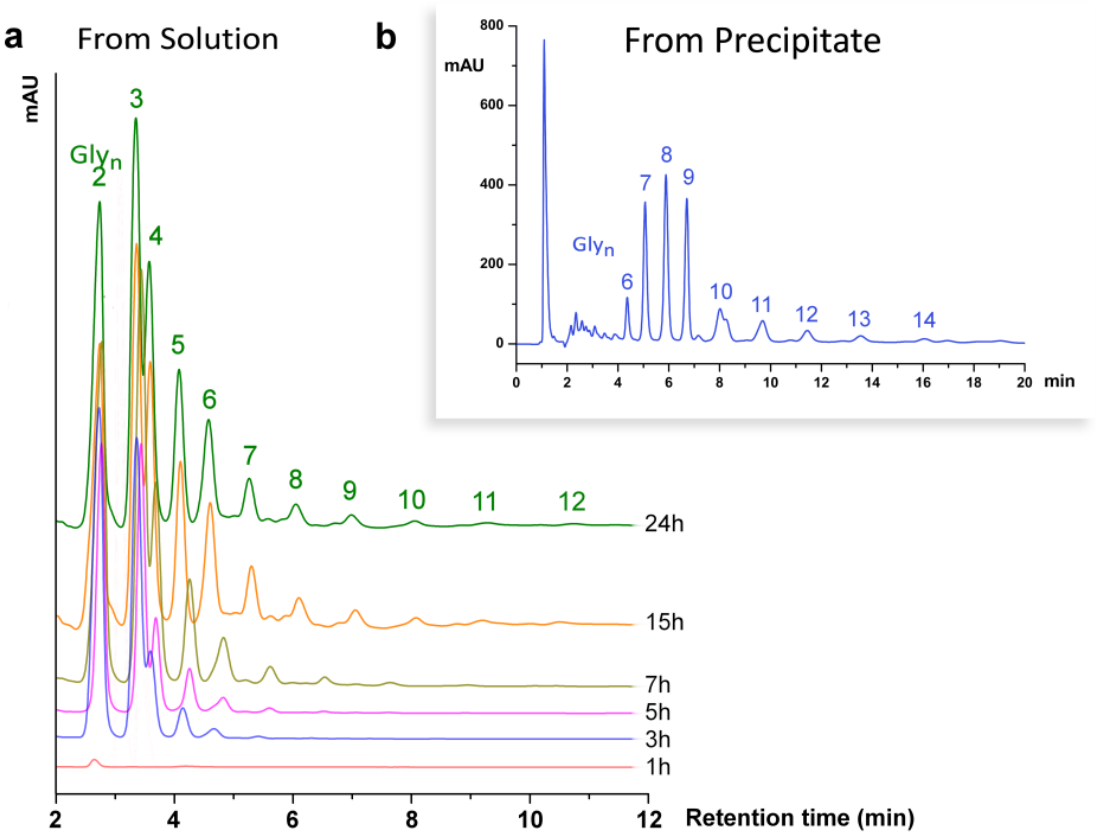


Figure 65: **a)** IP-HPLC chromatograms of a standard glycine condensation reaction, showing product oligomer length increasing with dehydration time (between 1 and 24 hours). **b)** Similar analysis of the solid precipitate formed during the reaction after two dehydration-hydration cycles, taken up in a 0.1% TFA/H_2O solution, showing longer (sparingly soluble) oligomers. In both **(a)** and **(b)**, Gly_n oligomer peaks are labelled with the corresponding value of *n*; assignment of the peaks was established by comparison to Gly standards and by separate MS confirmation.

Having demonstrated unactivated glycine oligomerisation, we used the APS to systematically investigate the influence of other variables such as temperature, pH, number of cycles and cycle duration on the distribution of the oligopeptides formed. On raising the temperature of reaction from 90 °C to 130 °C, we observed a general increase in yield (from less than 1% after 1 h at 90 °C to ca. 50% after 15 h at 130 °C, see Fig. 66a&67). However, later in reactions at higher temperatures, we also observed the appearance of a brown colour (from a colourless solution; and note that this coincides with an apparent drop in solution yields, see Fig. A6). This may be due to both precipitation of longer oligomer products or decomposition. Both are possible, but we note that no significant new peaks are resolved in IP-HPLC for decomposition products. Setting the pH of the amino acid input solutions was found to influence the reaction (see Fig. 66b&68): we observed a less than 0.1% yield at pH 6.1, rising to *ca.* 45% yield at pH 9.75 in more basic conditions, and to *ca.* 20% in more acidic conditions.

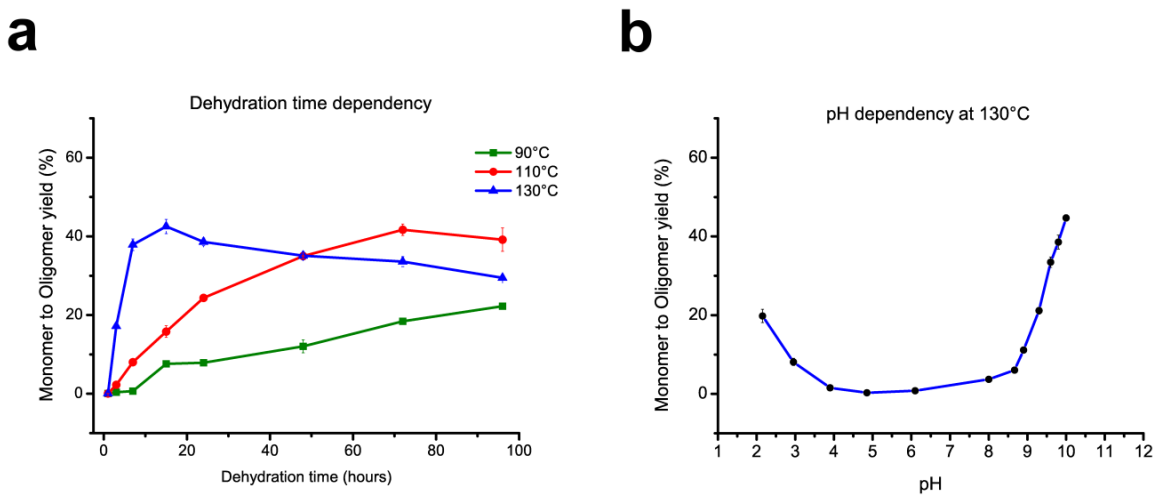


Figure 66: Graphs showing oligomer yield from a single cycle as a function of **(a)** dehydration time and **(b)** pH (at 130ºC for 24 h).

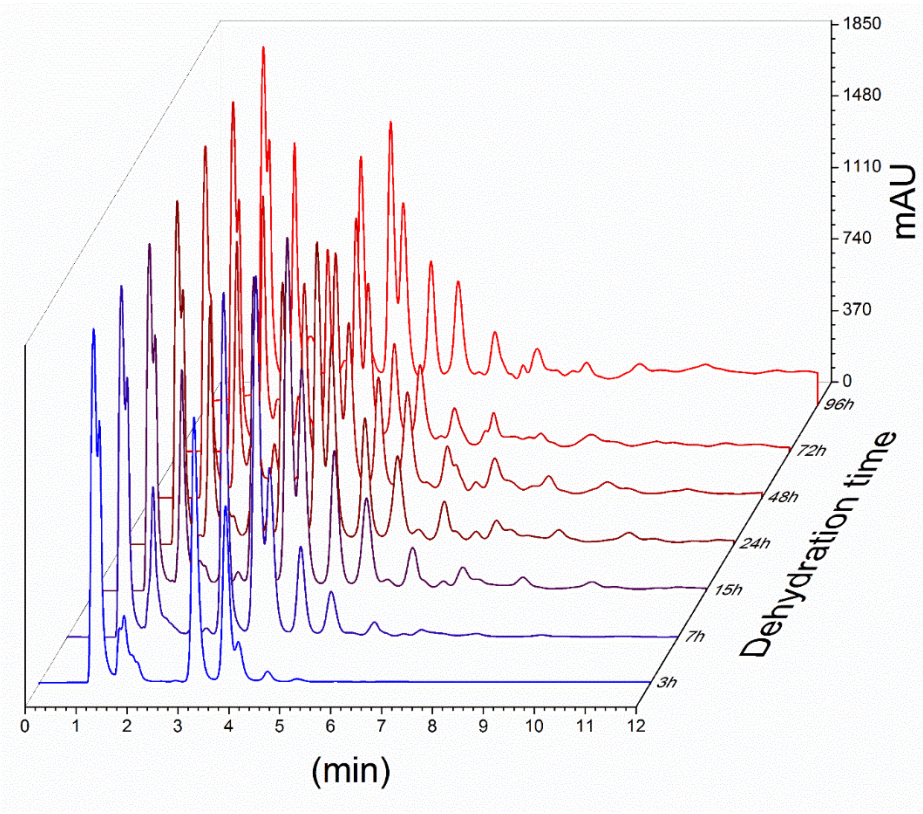


Figure 67: IP-HPLC chromatograms showing the oligomerisation products of glycine as a function of the dehydration time at130 ºC (pH=9.8).

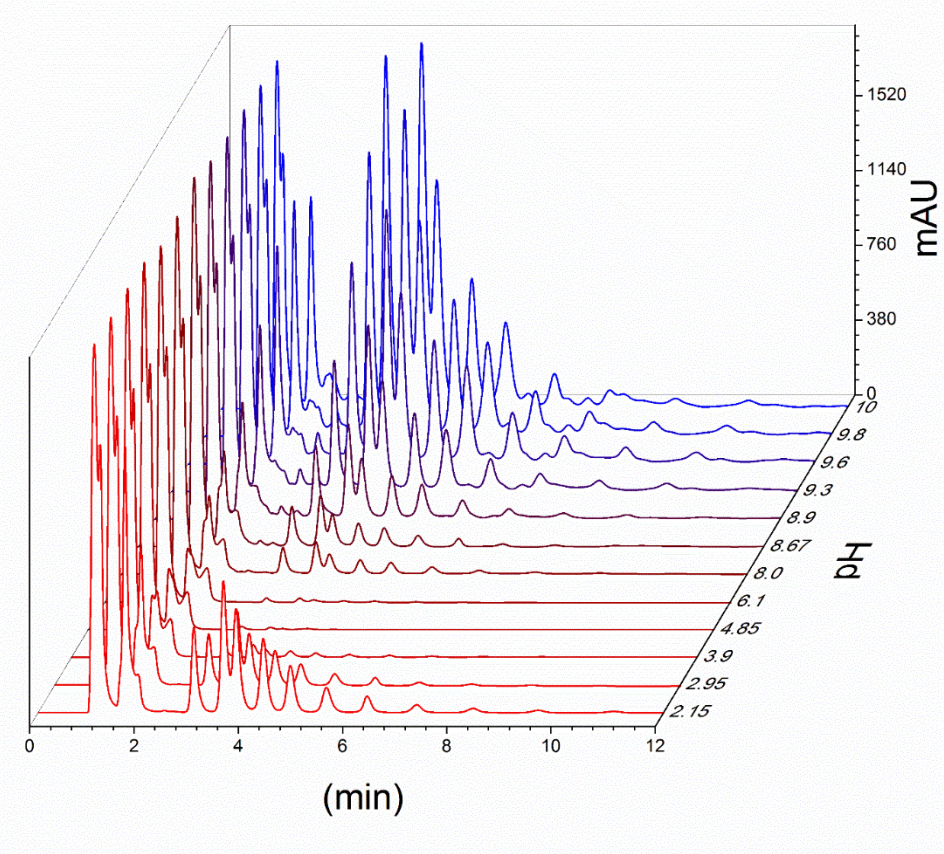


Figure 68: IP-HPLC chromatograms showing the oligomerisation products of glycine as a function of the pH in a single dehydration cycle (15h) at 130ºC.

As the number of cycles increased the distribution of oligomer chain-lengths observed in solution shifted. The yield of the lower oligomers decreased, while that of the longer oligomers in solution remained fixed and the amount of solid material precipitated increased (see Figs. 69&70).

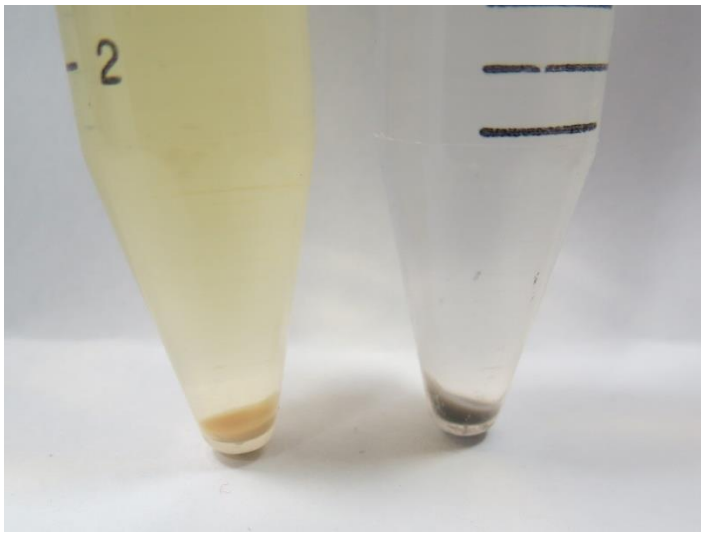


Figure 69: Insoluble product formation for a standard glycine condensation reaction (130ºC) under basic conditions (left) and acidic conditions (right), for a total number of 4 dehydration-hydration cycles (15h each).

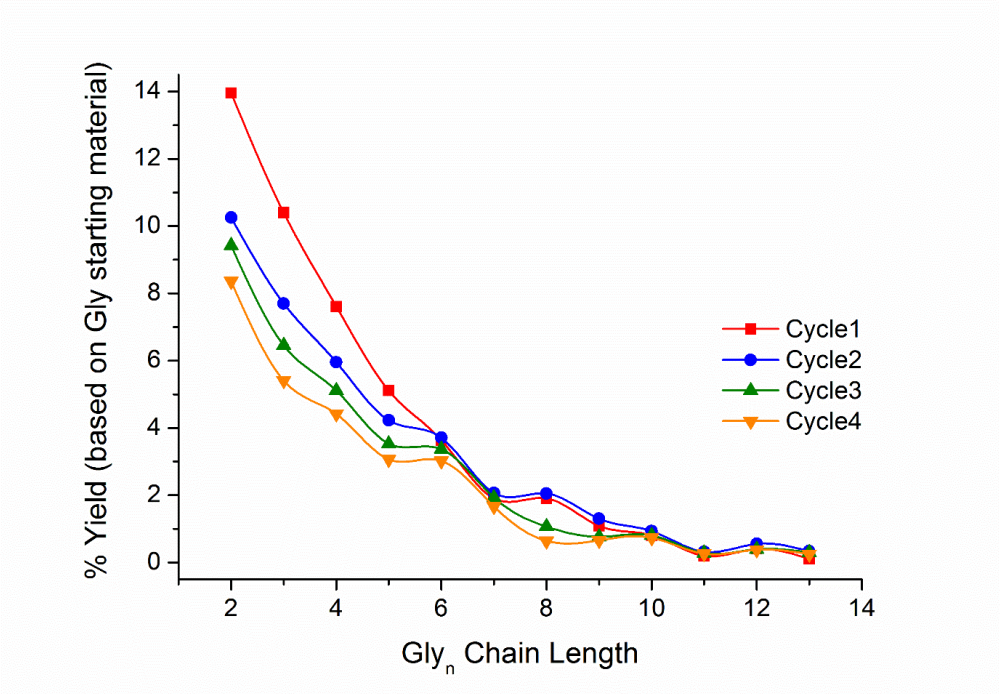


Figure 70: Distribution of Glyn chain lengths as a function of dehydration cycle number (oligomer yields expressed as % of initial Gly added).

The yields achieved by the process described above are considerably in excess of those previously reported to result from similar reactions:[95] this is likely the result of exploring the effects of several parameters at once, without the constraint of hypotheses on the nature of optimum conditions. While the system under consideration is more complex than might previously have been imagined, we note that the starting pH is of particular importance, with unprecedented yields observed from acidic and basic reaction mixtures. This is readily rationalised mechanistically: it has previously been shown that glycine dimerisation proceeds most readily at high pH where unprotonated amine groups are more nucleophilic and hence more readily attack partner carbonyls[105], whereas at neutral pH glycine monomers are zwitterionic and interactions between the charged amino group and a charged carboxyl group reduce reactivity.[105] The oligomerisation of glycine under acidic conditions has not attracted particular attention because of the poor nucleophilicity of protonated amines, however we note that acid catalysis should not be unexpected as the OH group of the carboxylate becomes a better leaving group (H_2O) on protonation and equilibria supplying a small amount of deprotonated amines should always be operative.

While our reaction conditions incorporate some NaCl (to maintain ionic strength at a similar order of magnitude in the absence of strong base to adjust pH), these conditions are very different to many previous studies (where often [NaCl] >> [Gly]).[95]

Furthermore, our initial observations for reaction of glycine suggest that NaCl can be omitted without significant loss of yield.

3.2.2.3 Making and breaking peptide bonds

In addition to the oligomerisation of glycine, we found that reaction of both glycinamide and DKP produce reasonable yields (>10%) at 130 °C, giving oligomers up to 10-mer (see Figs. 71 & 72). The role of DKP has been subject to debate: either seen as a 'dead end'/ thermodynamic 'sink', [220] or as able to react (often with caveats such as the presence of 'free' amines).[105] Observation of condensation of these amides, and of possible traces of glycine monomer in DKP reactions, leads us to pose an important question: are peptide bonds being concurrently made and broken during this process? To test this, we studied the reaction of linear Gly_2 dimers under the same conditions (see Fig.73). Along with the series of event-numbered Gly_x oligomers expected were only bond formation occurring, we observed the formation of an odd-numbered series (x = 3, 5, 7, etc.; see Fig. 73). This provides clear evidence that peptide bonds are both being formed, and being broken, concurrently raising the possibility of dynamic combinatorial processes.

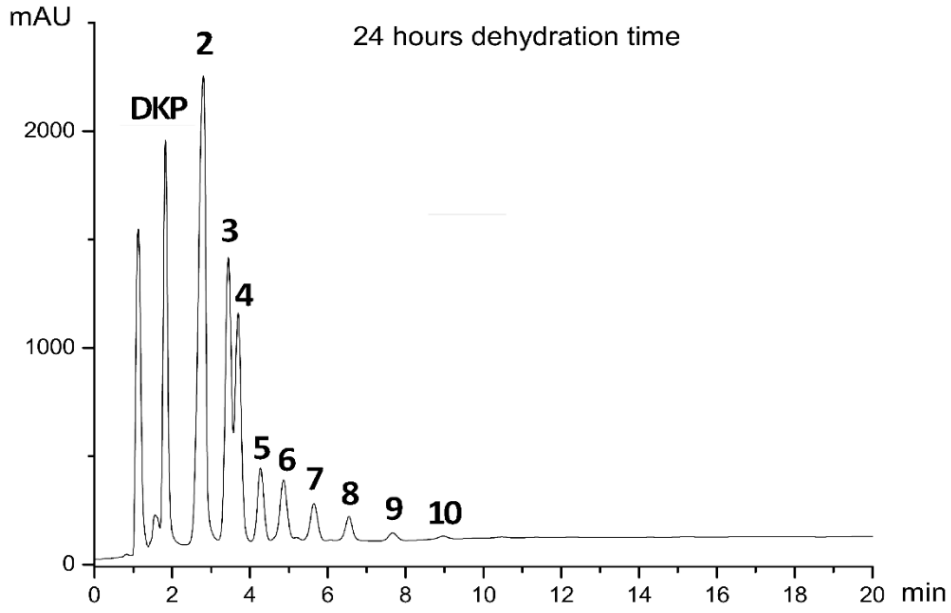


Figure 71: IP-HPLC chromatogram of DKP oligomerisation products (24h, 130ºC).

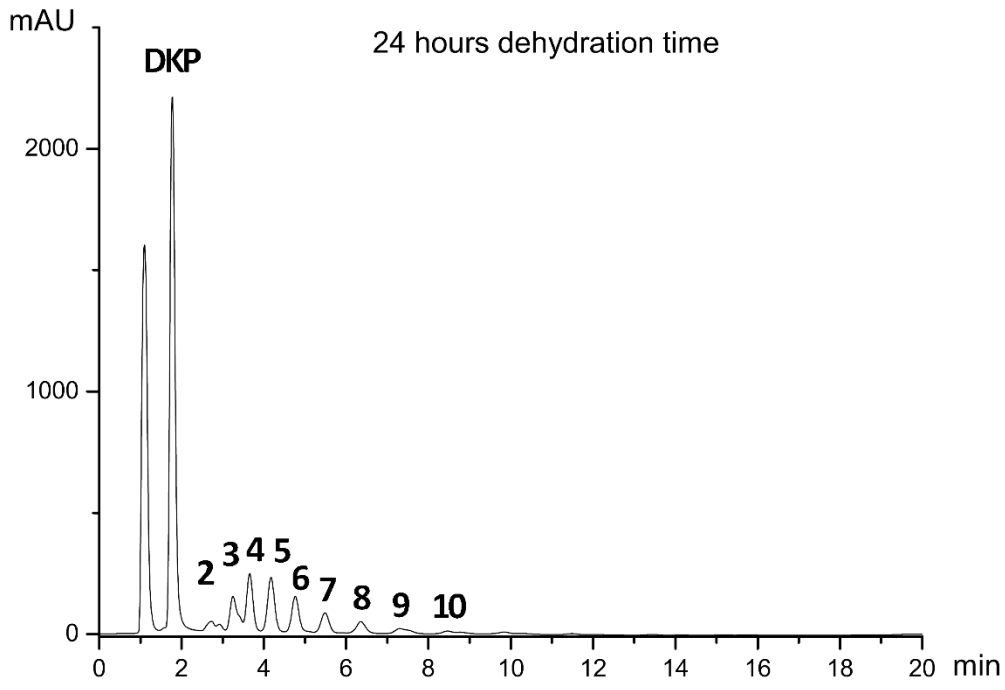


Figure 72: IP-HPLC chromatogram of glycinamide oligomerisation products (24h, 130ºC).

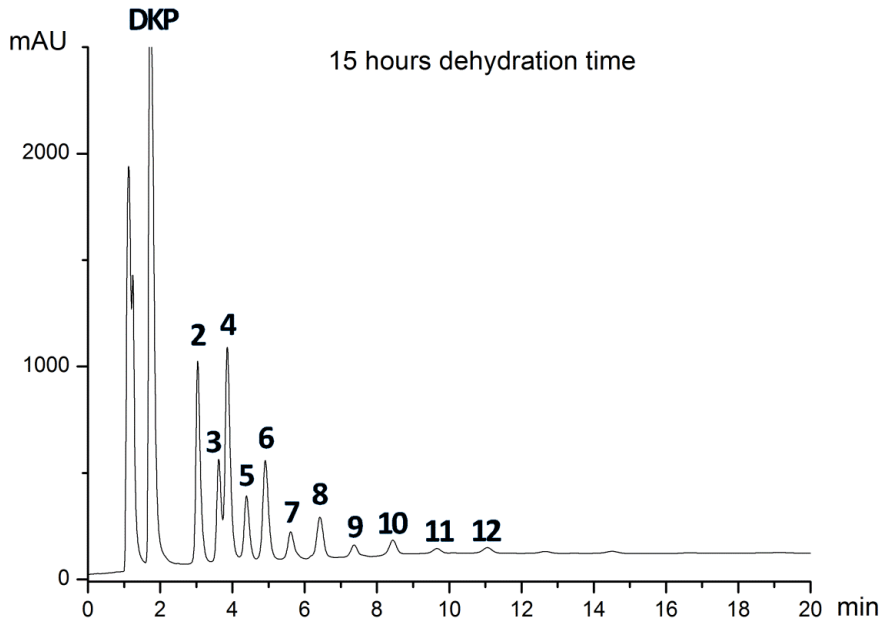


Figure 73: IP-HPLC chromatogram of of glycylglycine oligomerisation products (24h, 130ºC).

3.2.3 Exploring environmental conditions

3.2.3.1 NaCl dependency

In the previous section, we reported the formation of glycine oligomers under very simple conditions by studying different combinations of several parameters simultaneously. This first approach led us to find out which combination of conditions (pH, temperature, and dehydration time) contributed the most to achieve high yields of glycine oligomers. In all the reactions, the pH was adjusted with strong bases or strong acids. Since we wanted to maintain the ionic strength at a similar order of magnitude in all our experiments, we decided to include a small amount of NaCl.

Extensive previous work from Rode *et al.*[221] proposed that high concentrations of NaCl (>3M) can act as strong dehydrating agents even in aqueous solution. When a divalent metal ion (Cu^{+2}) and glycine was added to a highly concentrated salt solution, a Cu complex consisting of one central divalent copper ion, a chloride ligand, two amino acids and two axial water molecules was formed. The complex formation catalysed the condensation between the unprotonated amine from one amino acid and the carboxylic acid of the adjacent amino acid. In this way, glycine oligomers up to 3-mers could be formed in small yields in aqueous solution. Later work from Rode *et al.*[95] demonstrated the formation of glycine oligomers up to 6-mers (<1%) when a solution containing glycine, NaCl and $CuCl_2$ (0.1, 0.5 and 0.05M respectively) was dehydrated-rehydrated at 90 °C for 4 cycles. However, the role of NaCl in the SIPF reaction was to ensure the presence of chloride ligand and to provide a dehydrating environment.

As stated by Rode: "Na^+ is not involved in the reaction, except for the role of taking water into its hydration shell, which can be performed by other comparable ions as well" and that "coordination of Cl^- to Cu(II) is an essential feature of the reaction".[95]

Also, recent work by Deamer *et al.* [145] showed the improved formation of long RNA-like oligomers in the presence of different monovalent salts. In this particular case, it is suggested that the eutectic phases formed upon the dehydration of monovalent salt solutions helped to concentrate and organise mononucleotide monomers that then undergo condensation reactions to form oligomers.

When this work was reviewed for publication, it was suggested by the reviewers that the NaCl included in our reactions could have played a key role other than to maintain the ionic strength, such as to promote the formation of peptide bonds as in Rode's SIPF reaction.

Since the presence of NaCl seemed a very interesting variable to study, we decided to run a standard glycine oligomerisation reaction (pH=9.8) at different NaCl concentrations to discover whether it had an effect on the final reaction yields.

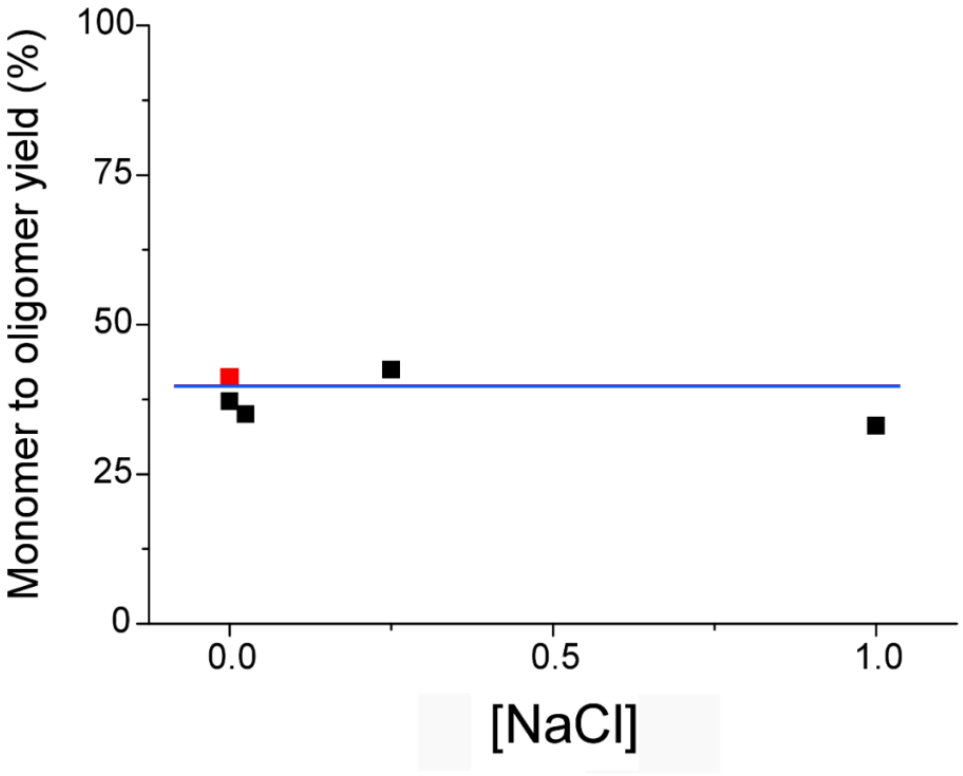


Figure 74: Monomer to Oligomer yield (%) vs. added [NaCl] (M). The red point indicates the yield obtained in an experiment ran under the same conditions, but using LiOH to adjust the pH in the absence of NaCl.

As observed in Figs. 74&75, the yield of glycine oligomerisation did not show any variances when the reaction was performed under different NaCl concentrations. Furthermore, although NaOH was used to adjust pH, reactions using LiOH instead show that Na^+ is not crucial. We note that a small concentration of alkali metal ions were always present to adjust the pH (as the simplest means less likely to get involved in the reaction). However, the Na^+ concentration resulting from adjusting the pH of our sample to 9.8 was 3.5 times lower than the initial glycine concentration in our reaction ($[Na^+]$=0.025M, [Gly]=0.0875M).

On the contrary, Rode's evaporation reactions (where NaCl concentration is 5 times greater than the initial glycine concentration, $[Na^+]$=0.5M, [Gly]=0.1M) produced no detectable species longer that tetramers where the yield was <1%.[95]

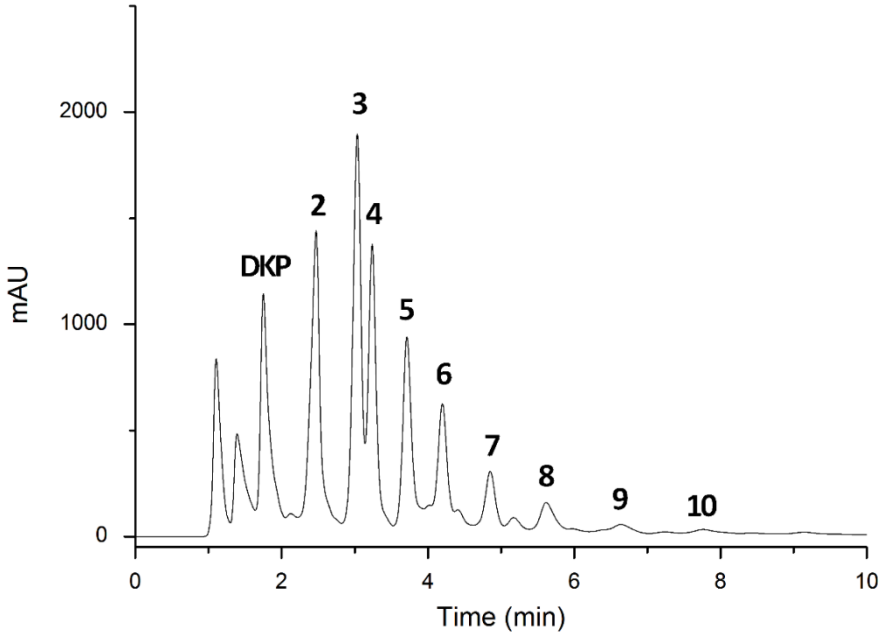


Figure 75: IP-HPLC chromatogram of glycine oligomerisation products corresponding to the black point at an added [NaCl] = 0 in the previous plot. In this sample, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was diluted in a total volume of 4ml. The pH was adjusted to 9.75 by adding 100 µl of a 1M NaOH solution. No NaCl was added.

In light of these results, we confirmed that NaCl could be omitted in this reaction for these specific set of conditions (pH=9.8, 130 °C, 15h dehydration time) without significant loss of yield. However, we decided to study the effect of adding NaCl at different pH values to discover whether NaCl did not contribute to the oligomerisation reaction rather than to maintain the ionic strength.

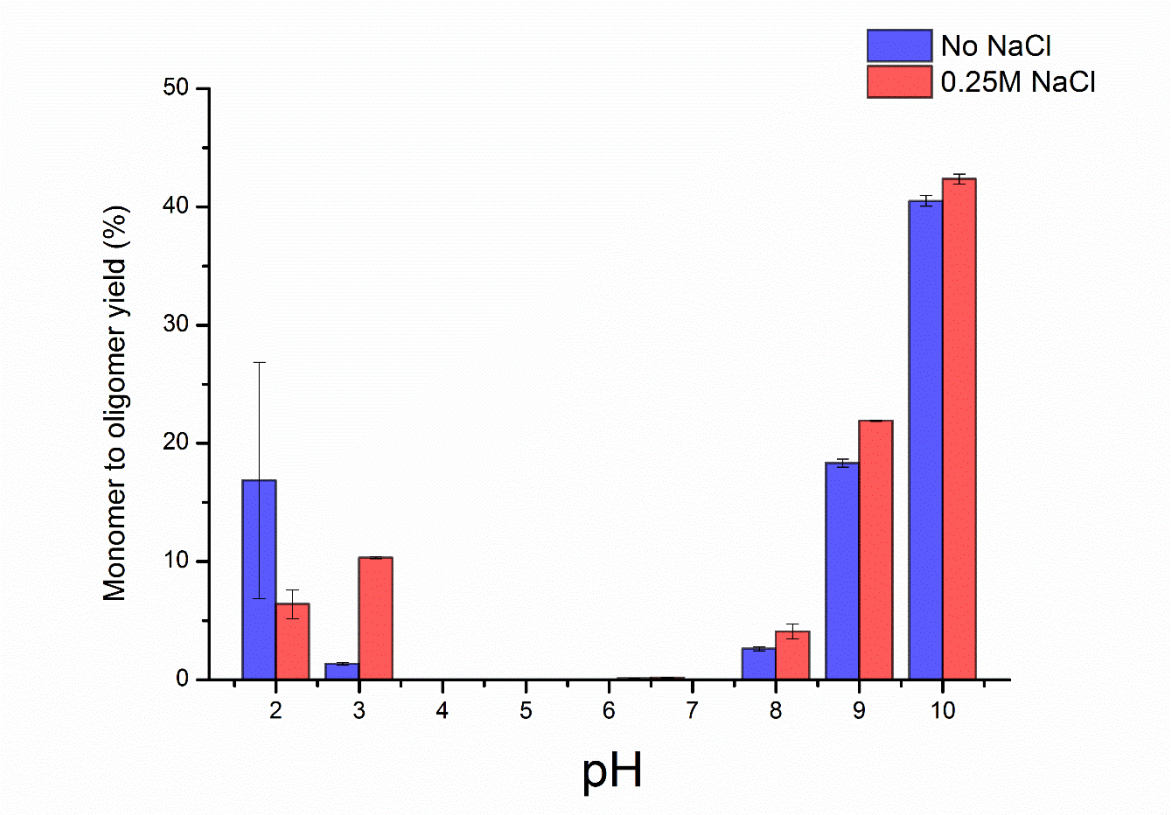


Figure 76: Monomer to oligomer yield (%) vs. pH in the absence of NaCl (blue) and using a 0.25M NaCl solution (red).

As observed in Fig. 76, only small differences in yield (<5%) resulted from the addition of NaCl under basic conditions. When the reaction was performed under acidic conditions (using HCl to adjust the pH), we observed a clear increase in yield when the reaction was performed in the presence NaCl at pH=3 (~10%). However, when the reaction was performed at pH=2, we could not conclude whether the presence of NaCl promoted the oligomerisation reaction due to large variances between experimental repeats. In our acidic reactions, water was removed by heating the reaction vial at 130 °C for 15h, leaving a solid. HCl is a gas at room temperature and high-concentrated HCl (38% kG HCl/Kg) solutions have a boiling point of approximately 50 °C.[222] As described in section 3.2.2.2, this reaction readily proceeds both under basic and acidic conditions. However, under the high temperature conditions of these reactions HCl would volatilise. In its absence, the low pH would not be maintained throughout the reaction resulting in inefficient polymerisation.

We noticed that when a lot of vials were placed at the same time in the oven, some of the reactions performed under acidic conditions produced very different yields, which explained the high error bars in some of the experiments.

More interestingly, some reactions run under what were thought to be the exact same conditions showed different colours when the reaction was finished. We then realised that when a large number of vials were placed in the convective-flow oven, since it is a partially-closed system, the relative humidity of the chamber increased during the first stages of the reaction, which led to a slower evaporation rate of the samples. Note that conclusions drawn from previous experiments performed under the same conditions would not change since they did not include a volatile acid, such as HCl.

In order to understand whether the rate of evaporation could have had an effect in these reactions, we performed standard glycine oligomerisation reactions at different evaporation rates.

3.2.3.2 Controlling the rate of evaporation

In this preliminary experiment, we wanted to study the oligomerisation of glycine monomer as a function of the evaporation rate. This was accomplished by performing two reactions in parallel that contained the exact same starting solution (4 ml, 0.0875 M glycine monomer) but within different containers: a 3D printed plastic lid with a 2.5 mm diameter hole was placed on one of the containers (slow evaporation rate), whilst the other container was left open (fast evaporation rate). Both reactions were left to dehydrate in the oven at 130 °C for 15h. The reaction was also performed at three different pH values (HCl and NaOH were used under acidic and basic pH, respectively).

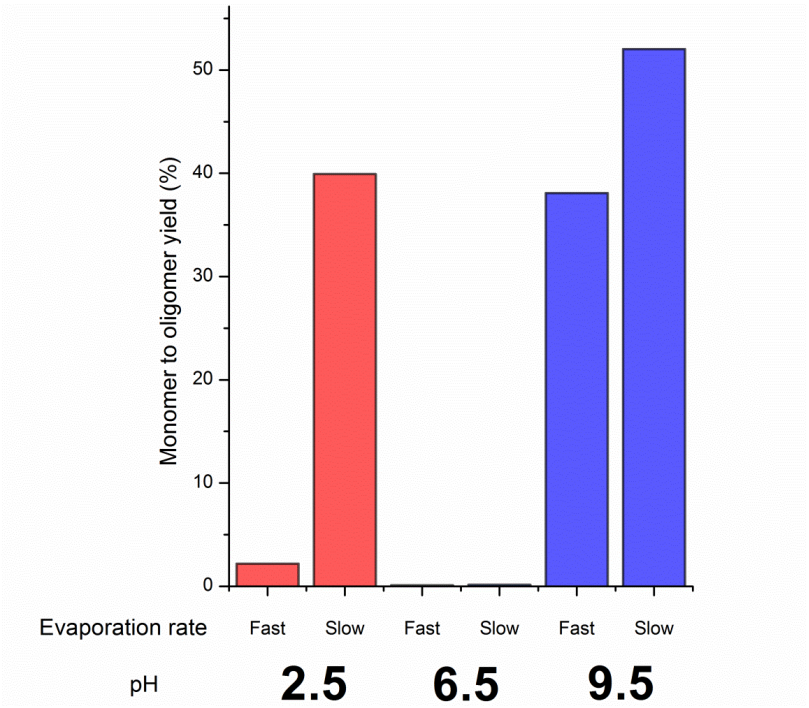


Figure 77: Monomer to Oligomer yield (%) vs. pH at two different evaporation rates.

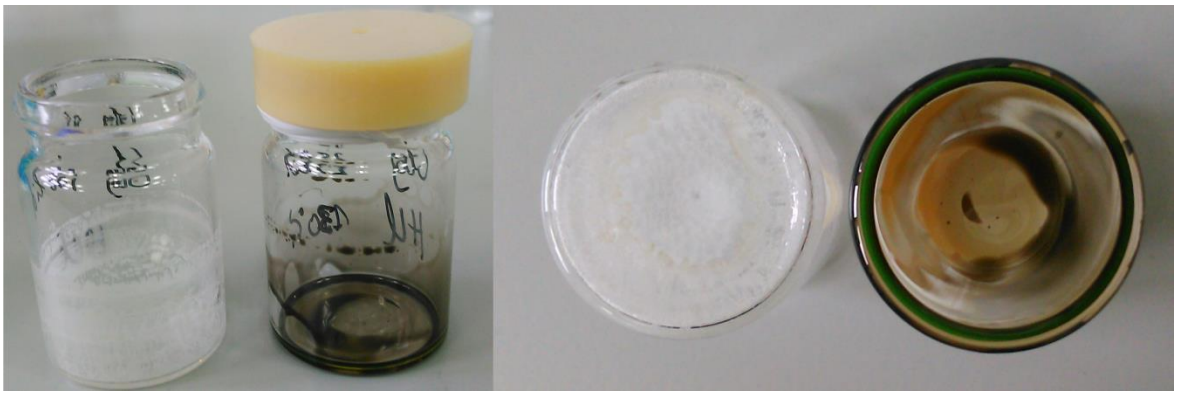


Figure 78: Two reaction vials containing the exact same starting solution (Gly 0.0875M, pH=2.5 adjusted with HCl), dehydrated for 15h at 130 °C at different evaporation rates. The vial on the left was left to dehydrate open, whilst a 3D-printed plastic lid with a 2.5mm hole was placed on the vial on the right.

Given that both reactions occur at different evaporation rates, whilst the experiment time is the same (15 h at 130 °C), the amount of time they reside in the absence of water is different. Note that in the experiment performed at a low evaporation rate it will take longer to remove the water by evaporation.

As observed in Fig. 77, a slow rate of evaporation dramatically increased the oligomerisation yield in the acidic reaction (where HCl was used to adjust the pH), from ca. 2% under fast evaporation, up to ca. 40% when a 3D-printed lid was placed on the vial. The same effect was observed in the basic reaction (where NaOH was used to adjust the pH) at lower extent.

We also noted a very different appearance between the two reactions run under acidic conditions (see Fig. 78). Whilst a white crystalline material was obtained when the reaction was performed under a fast evaporation rate, the reaction under slow evaporation conditions produced a darker solid.

Previous work by Viedma[223] studied whether evaporating solutions of amino acids at moderate temperature (100 °C) could either result in the formation of a crystalline phase or a non-crystalline solid phase depending on the experimental conditions. When a solution of amino acids becomes saturated by evaporation, crystallisation can take place. However, depending on a different number of conditions (temperature, evaporation rate, stirring, etc.), a saturated solution can remain without inducing crystallisation for some time. Under these conditions, where the water activity is low and the amino acid monomers are still able to react within a saturated solution, the condensation of amino acids into peptides could take place.

Our preliminary results indicated that by decreasing the evaporation rate we may prolong the time in which amino acids are able to react within a saturated solution at low water activity. Also, in the case of HCl, decreasing the rate of evaporation could also decrease HCl loss as a gas, thus extending the time at which the reaction is maintained at low pH.

Having demonstrated that the slow evaporation of our amino acid solutions had a great effect on the reaction yields, together with the pH, we then tested this for different acids and bases, to demonstrate that it was not specific to NaOH and HCl.

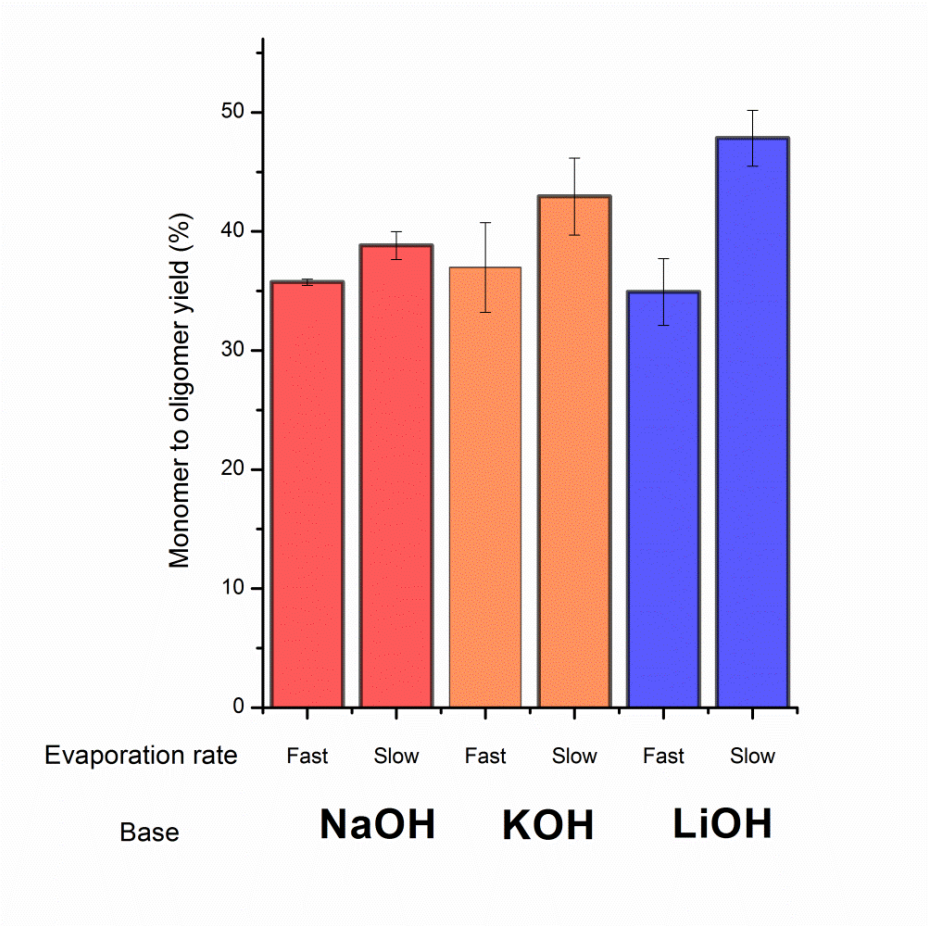


Figure 79: Monomer to Oligomer yield (%), using three different bases to adjust the pH (9.5) at two different evaporation rates.

When three different bases were used to adjust the pH to 9.5, we observed that performing the reaction at a slow evaporation rate produced higher yields (between 2% and 15%) than performing it at faster evaporation rates (Fig. 79).

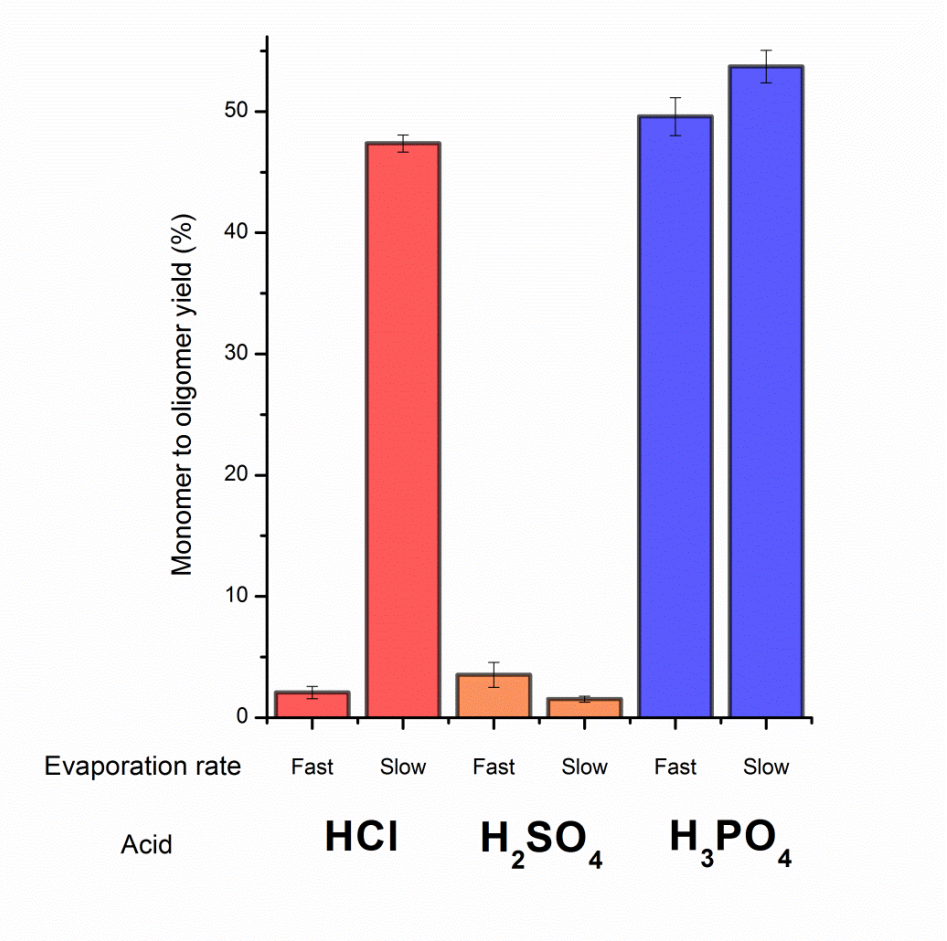


Figure 80: Monomer to Oligomer yield (%), using three different acids (HCl 1M=175µl, H_2SO_4 1M=100µl and H_3PO_4=225µl) to adjust the pH (2.5) at two different evaporation rates.

As observed in Fig. 80, very different results were obtained when the same reactions were performed by using three different acids. We have corroborated that in the case of HCl, slowing the evaporation rate dramatically increases the oligomerisation yield.

We did not observe an increase in yield when the reaction was performed at a slow evaporation rate in the presence of H_2SO_4, and the overall yields were significantly lower compared to the other acids employed. On the contrary, oligomerisation did proceed efficiently in the presence of phosphoric acid both under fast and slow evaporation rate conditions, with a slight increase when it was performed under slow evaporation conditions.

Overall, we have observed that slowing the evaporation rate results in an increase in yield for this particular reaction, except when H_2SO_4 was used to adjust the pH.

Whilst further experiments would be required to fully understand the nature of these results, overall, we have observed that slowing the evaporation rate results in an increase in yield for this particular glycine oligomerisation reaction, except when H_2SO_4 was used to adjust the pH.

In this set of experiments, different amounts of HCl, H_2SO_4 and H_3PO_4 were used to adjust the pH of the samples to 2.5 (175µl of a 1M HCl solution, 225µl of a 1M H_3PO_4 solution and 100µl of a 1M H_2SO_4 solution were added to 3.5ml of a 0.1M glycine solution, respectively).

As opposed to HCl, the boiling points of H_2SO_4 and H_3PO_4 are over the temperature in which the reactions are carried out, meaning that they remain as solvents in the vial when the water has been completely removed by heating. Since different amounts of H_2SO_4 and H_3PO_4 were used to adjust the pH, a different volume of acid would be present in the reaction after the water has been evaporated (11µl of H_3PO_4 and 5µl of H_2SO_4). Both H_2SO_4 and H_3PO_4 are strong dehydrating agents and can act as solvents in the absence of water, hence, we suspect that the reaction outcome could change if different volumes are present after complete evaporation.

3.2.3.3 Exploring the differences between H_2SO_4 and H_3PO_4.

In our preliminary work, we just studied the effect of different parameters (temperature, dehydration time, and pH) on a standard glycine oligomerisation reaction. The remarkable differences in yield between H_2SO_4 and H_3PO_4 led us to believe that other parameters, such as the volume of acid present upon dehydration, could play an important role on this reaction.

For this reason, we decided to perform a series of experiments were different amounts of H_3PO_4 and H_2SO_4 were added in order to understand their different behaviour. Initially, we increased the amount of acid in our reaction to observe if a higher amount of initial added acid would promote the formation of peptide bonds.

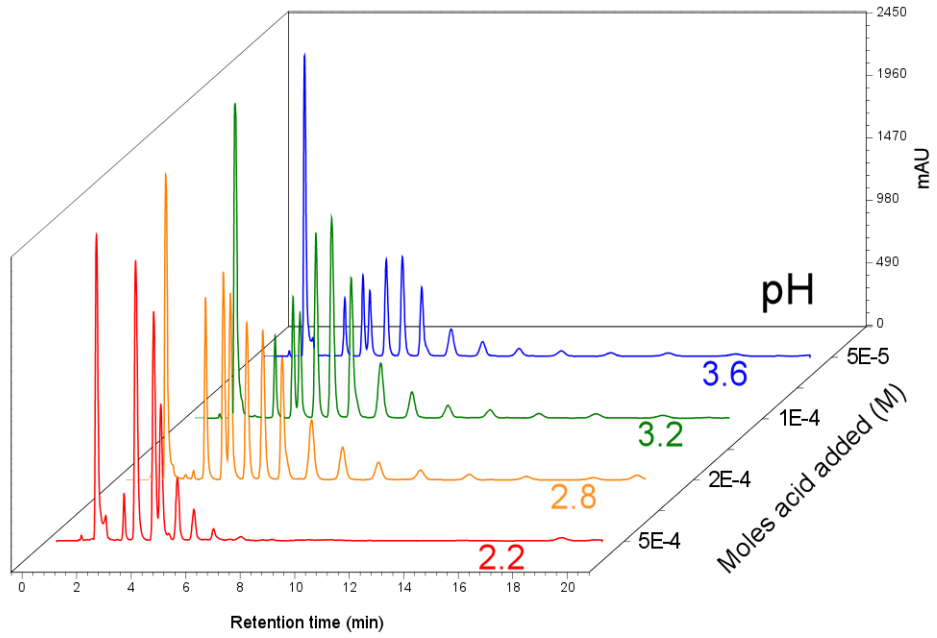


Figure 81: IP-HPLC chromatograms showing the oligomerisation products of glycine as a function of the number of moles of H_3PO_4 added in a single dehydration cycle (15h) at 130ºC.

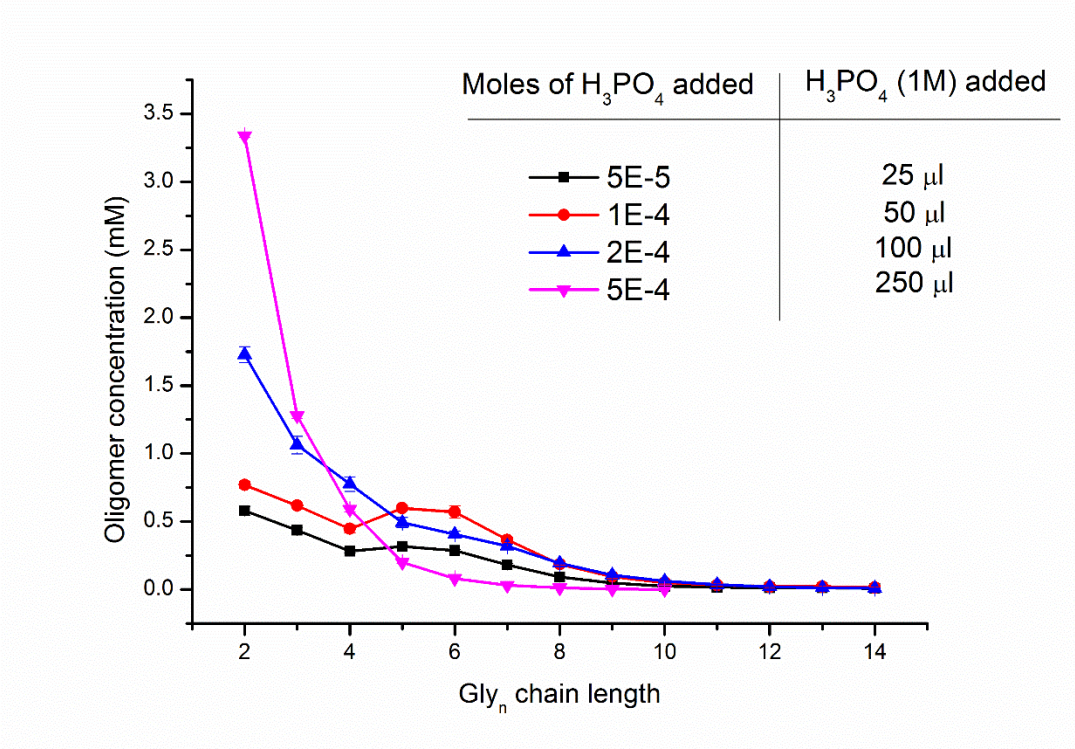


Figure 82: Distribution of Gly_n oligomer concentration as a function of the number of moles of H_3PO_4 added in a single dehydration cycle (15h) at 130ºC.

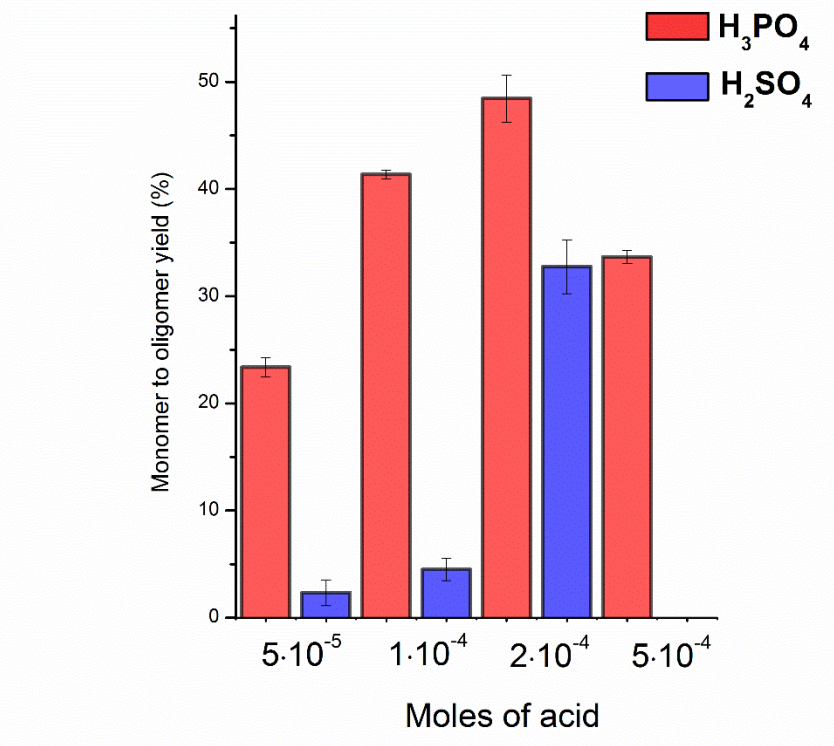


Figure 83: Monomer to Oligomer yield (%) as a function of the number of moles of H_2SO_4 or H_3PO_4 added in a single dehydration cycle (15h) at 130ºC.

.

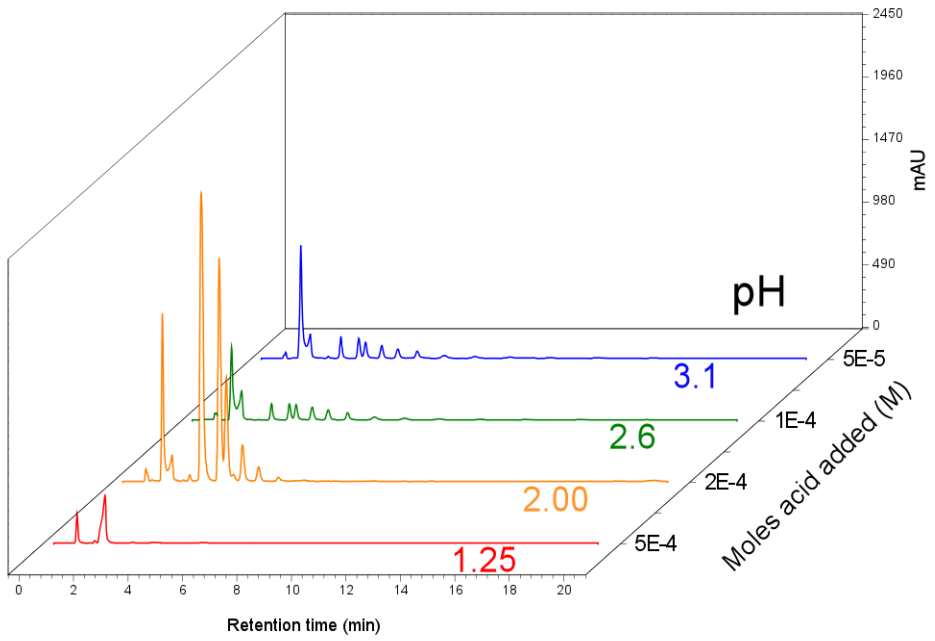


Figure 84: IP-HPLC chromatograms showing the oligomerisation products of glycine as a function of the number of moles of H_2SO_4 added in a single dehydration cycle (15h) at 130ºC.

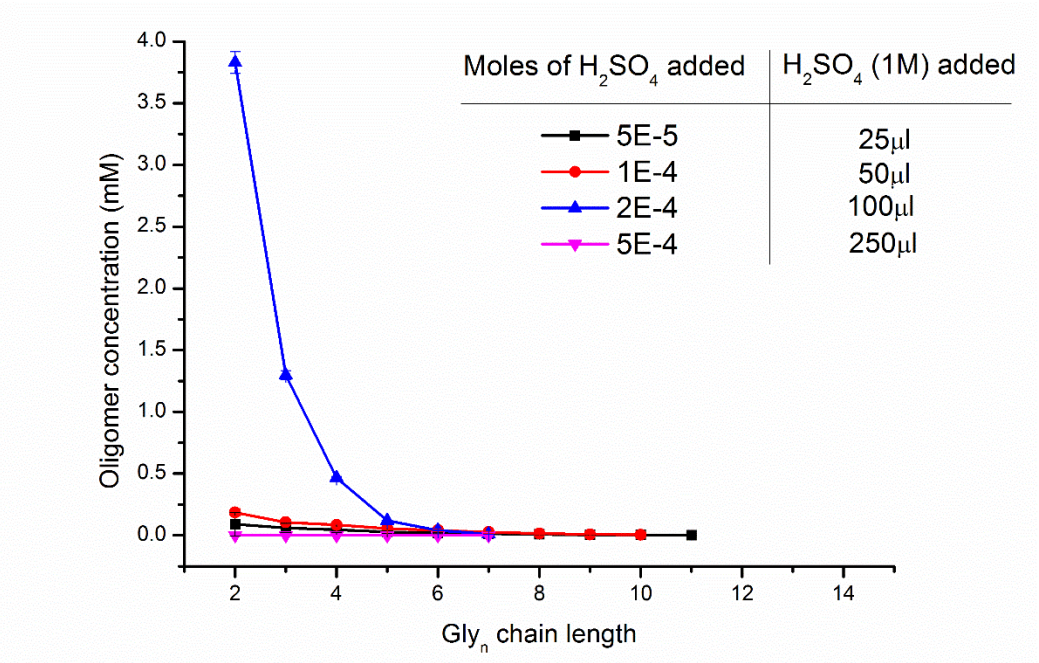


Figure 85: Distribution of Gly_n oligomer concentration as a function of the number of moles of H_2SO_4 added in a single dehydration cycle (15h) at 130ºC.

When the same set of reactions was performed in the presence of sulphuric acid, a similar trend was observed (see Figs. 84&85): longer oligomers were observed when a lower amount of acid was added, whilst a majority of short oligomers were obtained when the acid concentration was higher.

However, in this particular case, we noted that the reaction in the presence of sulphuric acid does not proceed as efficiently as in the presence of phosphoric acid. The longer oligomer concentrations (where n=5-10) are significantly higher in the presence of H_3PO_4. Also, the monomer to oligomer reaction yields are significantly lower compared to the ones obtained with H_3PO_4 (see Fig. 83). More interestingly, when a larger amount of H_2SO_4 was added ($5 \cdot 10^{-4}$ moles), we did not observe the formation of glycine oligomers.

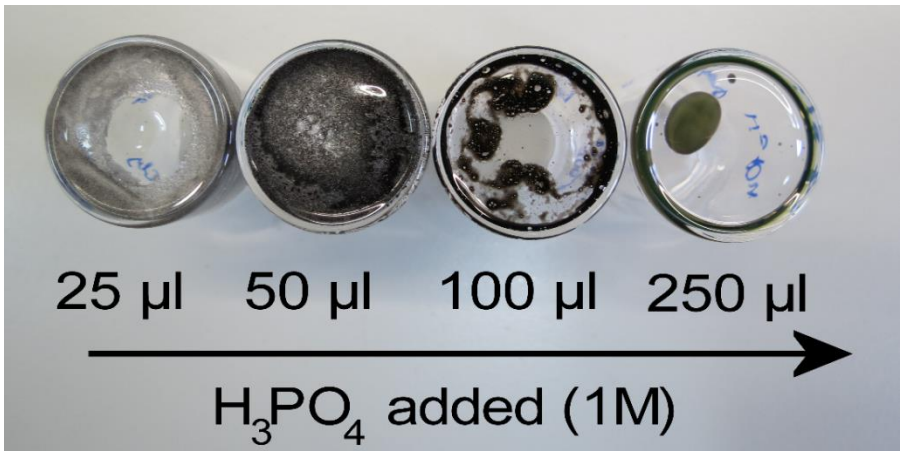


Figure 86: Dehydrated samples obtained when a different number of moles of H_3PO_4 were added (left: $5 \cdot 10^{-5}$ moles of H_3PO_4, to right: $5 \cdot 10^{-4}$ moles of H_3PO_4).

When observing the dehydrated samples after the reaction was finished, we noticed significant differences in their physical appearance. In the reactions performed with different amounts of H_3PO_4, a gradual browning of the dehydrated material became evident as the amount of acid increased (Fig. 86). Also, whilst the samples containing lower amounts of H_3PO_4 had a crystalline appearance, the sample with the highest concentration of acid had a non-crystalline appearance.

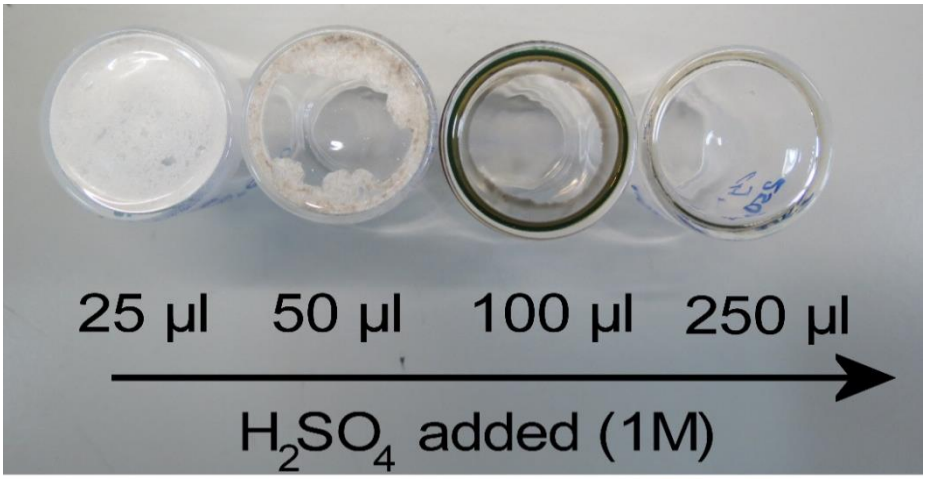


Figure 87: Dehydrated samples obtained when a different number of moles of H_2SO_4 were added (left: $5 \cdot 10^{-5}$ moles of H_2SO_4, to right: $5 \cdot 10^{-4}$ moles of H_2SO_4).

As observed in Fig. 87, the samples run with different amounts of sulphuric acid also had quite diverse appearances. The first two samples with the lowest amount of H_2SO_4 presented a crystalline appearance. The following sample (where $2 \cdot 10^{-4}$ moles of H_2SO_4 were added) resulted in the formation of a non-crystalline material. However, the sample with the highest concentration of H_2SO_4 was a clear liquid.

In light of these results, there are four main observations that can be made:

a) When adding the same number of moles of acid, H_3PO_4 promotes the formation of longer glycine oligomers compared to H_2SO_4.
b) In both cases, larger amounts of acid promote the formation of shorter oligomers, which could be attributed to a higher degree of peptide bond hydrolysis.
c) In both cases, when a larger amount of acid is added, the dehydrated sample has a non-crystalline appearance.
d) The final reaction yields initially increase when a higher concentration of acid is added. However, if too much acid is added, this results in a drop in yield.

Given that both H_3PO_4 and H_2SO_4 are liquids at the reaction temperature, they can serve as a solvent upon dehydration. On that basis, we expected to find an optimum solute to solvent ratio in which the formation of peptide bonds was enhanced due to a higher diffusion of the molecules within the solvent. We initially imagined that by adding an increasing amount of acid, we would obtain a higher reaction yield and the formation of longer oligomers. However, IP-HPLC analysis (see Figs. 82&85) clearly shows that shorter oligomers are obtained when a higher amount of acid is added. As mentioned before, a thorough study would be required to understand the rate of bond formation versus bond breaking in this particular system. Also, we observe that H_3PO_4 promotes the formation of longer oligomeric species compared to H_2SO_4.

However, what could explain the differences observed between these two acids?

Whilst the same number of moles of acid was added in the sets of equivalent reactions, (see Figs. 81&84) initial different pHs were obtained due to the different acid dissociation constants between H_3PO_4 and H_2SO_4 (when $2 \cdot 10^{-4}$ moles of each acid were added, the pH obtained was 2.8 for H_3PO_4, and 2.00 for H_2SO_4). Given that we initially observed an increase in the monomer to oligomer yield upon the addition of both H_3PO_4 and H_2SO_4, we assume that the addition of acid catalyses the formation of peptide bonds. Nonetheless, the different behaviours observed between H_3PO_4 and H_2SO_4 could be explained by their role in acid catalysis. There are two types of acid catalysis: specific acid catalysis and general acid catalysis.

Specific acid catalysis involves a rapid protonation of the substrate (the amino acid in this case), which then becomes more reactive. The protonation is so fast that it does not become the rate determining step.

This catalysis depends on the protonating power of the solution, so the catalyst must be a strong enough acid. Also, specific acid catalysis works only if the pH is similar to, or below the pKa of the conjugate acid of the substrate.

On the other hand, general acid catalysis involves transfer of a proton from a weak acid during the rate determining step. It is effective at neutral pHs, even above the pKa of the conjugate acid of the substrate.[224]

Given the difference in pKa values between H_3PO_4 (2.12, 7.21 and 12.67) and H_2SO_4 (-3, 1.99), H_2SO_4 would have more protonating power in solution, hence making it more suitable for specific acid catalysis than H_3PO_4. On the other hand, at a given pH (i.e. 2.5), H_3PO_4 would have only lost one proton, and it still would exist as its conjugate base ($H_2PO_4^-$). Therefore, H_3PO_4 could operate as a general acid catalyst, which would explain the different results obtained by using these two different acids.

As previously mentioned, bonds are concurrently being formed and broken during these dehydration reactions. It was observed that the monomer to oligomer yield initially increased as the concentration of acid was increased. However, the yield decreased if too much acid was added (Fig.83). The role of the acid as a catalyst could be both in the formation and hydrolysis of peptide bonds, and an equilibrium between these two processes could operate as a function of the concentration of acid presents and their protonating abilities.

However, considering that the reaction takes places when the water has been evaporated, the initial pH in the amino acid solution might not be relevant to the reaction outcome. Instead, the differences between H_2SO_4 and H_3PO_4 in their pure form, such as their dehydrating properties or their protonating abilities in the absence of water, might explain the different results obtained.

3.2.4 Section summary

Whilst we used an automated reaction system to discover and optimise the initial conditions for dehydration-hydration driven peptide bond formation, this simple reaction did not require automation. At its most straightforward, the reaction of a 0.0875 M aqueous solution of glycine monomers (adjusted to either basic or acidic pH) over a period of 15 hours at 130 °C in a simple glass vial was observed to produce oligomers of a length greater than the minimum previously shown to yield function ($n \geq 7$).[225]

Likewise, it is worth mentioning that this process, involving reaction of an unadulterated solution of glycine in water over natrolite mineral also gave peptide-oligomers up to 8-mer in only one cycle (Fig. 88).

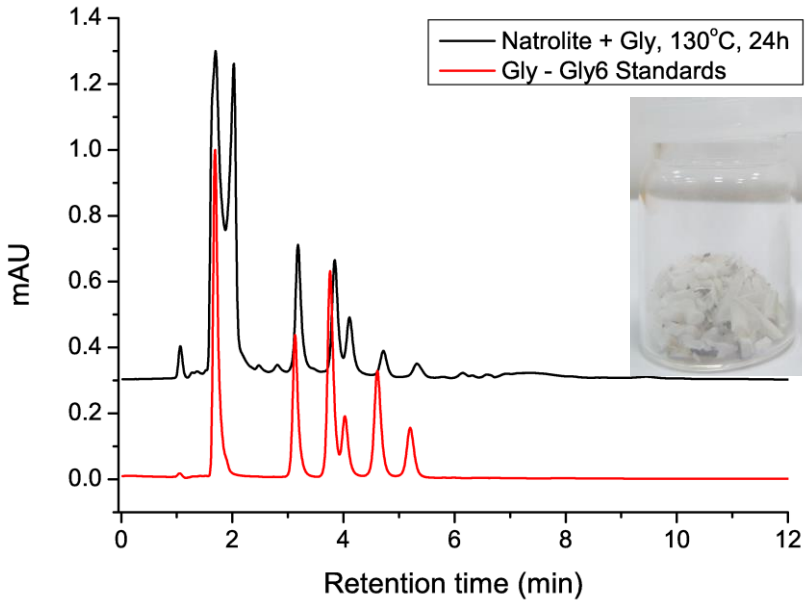


Figure 88: IP-HPLC chromatogram of glycine oligomerisation products with natrolite mineral (24h, 130ºC) compared to Gly, $(Gly)_{2\text{-}6}$ standards. Inset: a vial containing natrolite.

These conditions are similar to the formation of 'proteinoids' reported by Fox and coworkers.[81,35] Whilst previous authors reported formation of large insoluble products by such approaches[78,79,], it has been suggested that peptide bonds may have not been present in these products in great amounts[88,221], however in light of our results it appears that such bond formation may have been possible. Such a simple procedure should be of interest to both synthetic chemists and to those interested in how recursive chemistry could allow the gradual emergence of order without human intervention.[59]

In this study, our investigation of the reaction input and process variables revealed that peptide bond formation from unactivated amino acids is less challenging than previously imagined. This process is both simple and general, does not require catalysts or activating reagents, and can produce large yields of oligomers, for the majority of which $n > 3$. Preliminary tests suggest that a wide range of amino acids may be co-condensed in a similar manner. In the next chapter, we describe how we used this approach to investigate the propensity of small biases (intrinsic reactivity) to influence the structures and sequences resulting from reactions in fluctuating conditions. These aspects, not typically addressed when reporting uncontrolled peptide synthesis, are vital if such reactions are to be of use, whether for synthesis or the understanding of prebiotic peptide formation.

3.3 Hetero-oligomer synthesis exploration

3.3.1 Heteropolymers and expansion to other amino acids

Having established the formation of glycine homo-oligopeptides, we made some exploration of the co-condensation of glycine with several other amino acids (Ala, Asp, Glu, His, Lys, Pro, Thr, and Val). We initially used IP-HPLC to separate and quantify the oligomers produced in a glycine homo-oligomerisation reaction. The co-oligomerisation between different amino acids generated a much greater number of possible products that could not be resolved by this technique. HPLC-MS was used to separate and identify the co-oligomerisation products in our reactions. Furthermore, while not being formally quantitative, HPLC-MS can be used to gain a general understanding of trends of product distributions. An analytical workflow was developed to identify and measure the ion intensities corresponding to oligomeric peptides species.

In all of these cases RP-HPLC-MS analysis revealed the presence of many species consistent with $(Gly\text{-}X)_n$ hetero-oligomers and, furthermore, MS/MS analysis of representative masses (consistent with Gly_2X_2 tetramers) yielded α- and y-series fragments predicted by theory for such peptide structures (see Figs. A17-A32 for further details). Furthermore, combination of Gly, Ala, and Lys, also resulted similarly in apparent formation of hetero-oligomers containing residues of all three amino acids (see Figs. A33&A34). We note that the amino acids included in this preliminary investigation of co-oligomerisation incorporate a wide range of functionality in their side chains: carboxylic acid (Glu, Asp), primary (Lys), secondary (Pro) and aromatic (His) amines, alcohol (Thr) and hydrophobic (Ala, Val) groups. While fragments corresponding to linear peptide structures were observed, we also note that both branched structures and other forms of bonds (e.g. ester) may have been formed: indeed, given the presence of reactive side chains, this is not unlikely.

Figs. 89&90 show an example of the analytical workflow employed to identify and confirm the presence of peptide hetero-oligomers in our co-oligomerisation reactions.

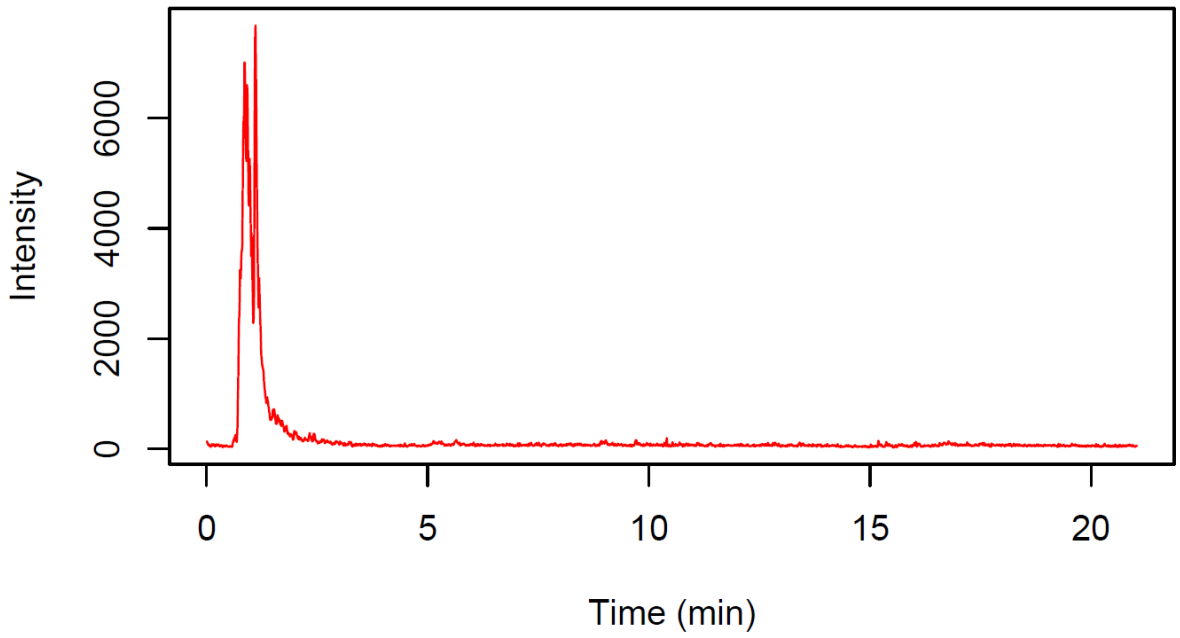


Figure 89: 'Virtual' BPC traces illustrating co-condensation of G and A monomers.

A BPC (Base peak chromatogram) is a chromatogram obtained by plotting the signal of the ions represented by the base peak detected in each of a series of mass spectra recorded as a function of retention time[226], often referred as the MS 'silhouette'. However, as shown in Fig. 89, we represented a 'virtual' BPC, which instead of representing all the ions analysed, it was constructed from the EICs observed for a list of possible oligomers incorporating both of the amino acids monomers present in our reaction.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
90.054	GGAA	A	[y1]1+	90.055	-0.001
90.054	GAGA	A	[y1]1+	90.055	-0.001
90.054	AGGA	A	[y1]1+	90.055	-0.001
115.05	GGAA	GG	[b2]1+	115.05	0
129.0656	GAGA	GA	[b2]1+	129.066	0
129.0656	GAAG	GA	[b2]1+	129.066	0
129.0656	AGGA	AG	[b2]1+	129.066	0
129.0656	AGAG	AG	[b2]1+	129.066	0
133.0613	AAGG	GG	[y2]1+	133.061	0
147.0752	GAGA	GA	[y2]1+	147.076	-0.001
147.0752	GAAG	AG	[y2]1+	147.076	-0.001
147.0752	AGGA	GA	[y2]1+	147.076	-0.001
147.0752	AGAG	AG	[y2]1+	147.076	-0.001
186.0859	GGAA	GGA	[b3]1+	186.087	-0.001
186.0859	GAGA	GAG	[b3]1+	186.087	-0.001
186.0859	AGGA	AGG	[b3]1+	186.087	-0.001
204.0965	AGGA	GGA	[y3]1+	204.098	-0.002
204.0965	AGAG	GAG	[y3]1+	204.098	-0.002
204.0965	AAGG	AGG	[y3]1+	204.098	-0.002

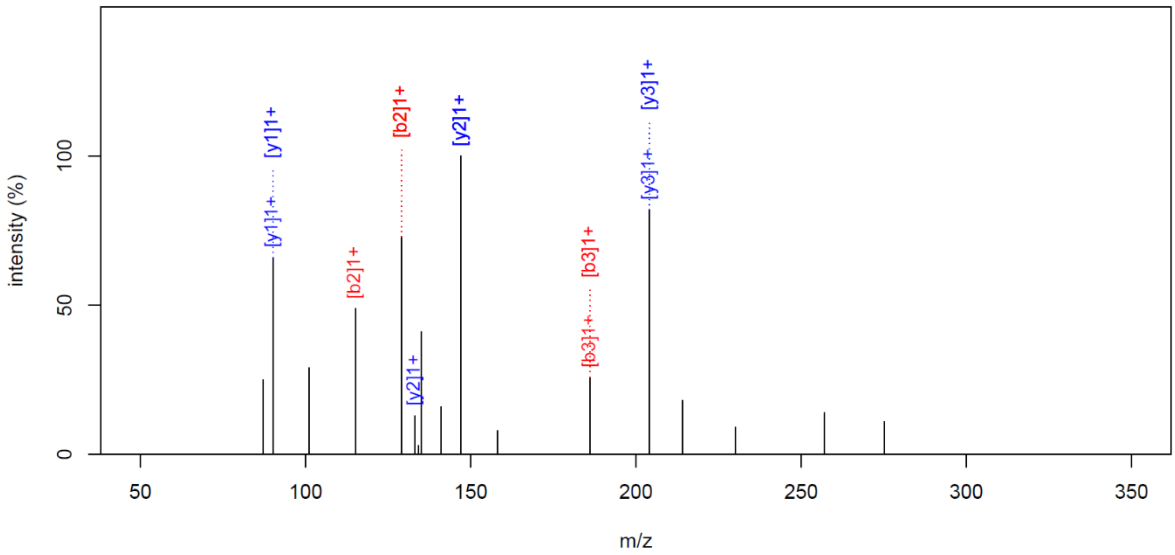


Figure 90: Example MS/MS analyses of products of G and A monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 275.135 (corresponding to a tetramer incorporating 2 x G and 2x A). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x A (ie. more than one sequence of the same composition is present, all incorporating G and A).

3.3.2 Exploring the differences between using H_2SO_4 and H_3PO_4 in co-oligomerisation reactions.

As explained in section 3.2.3.3, the use of different acids to adjust the pH resulted in differences in yield and oligomer distribution for a glycine homo-oligomerisation reaction.

We note that all these observations corresponded to a model reaction, in which a solution of glycine was the only starting material. Since we demonstrated that glycine co-oligomerises with a series of different amino acids, we decided to study the effect of using H_3PO_4 or H_2SO_4 in a reaction containing glycine and valine as starting materials.

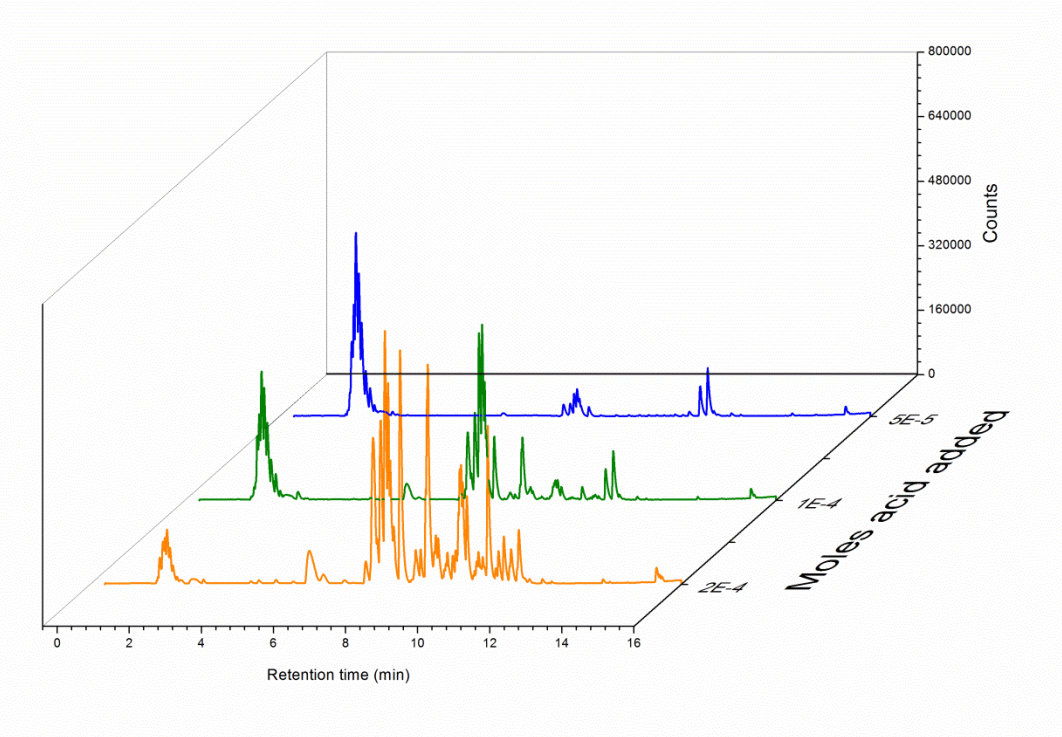


Figure 91: BPCs showing the co-oligomerisation products of glycine-valine reaction as a function of the number of moles of H_3PO_4 added in a single dehydration cycle (15h) at 130ºC.

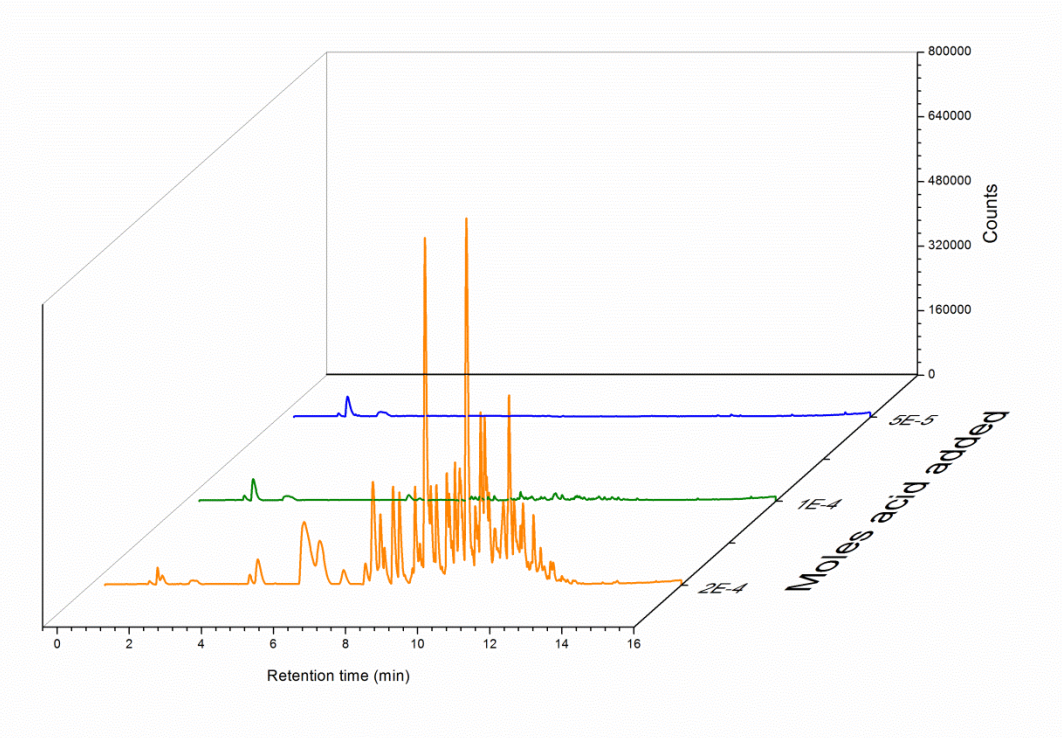


Figure 92: BPCs showing the co-oligomerisation products of glycine-valine reaction as a function of the number of moles of H_2SO_4 added in a single dehydration cycle (15h) at 130ºC.

Fig. 91 shows the BPCs obtained from a Gly-Val co-oligomerisation reaction upon the addition of different amounts of H_3PO_4. After integration of the peaks in the base peak chromatogram, we observed that the peaks between 2 and 3 min corresponded to oligoglycine species, whilst the peaks resolved between 8 and 12 min corresponded to glycine-valine co-oligomers. When $5 \cdot 10^{-4}$ moles of H_3PO_4 were added, we observed that oligoglycine species were the major product. However, an increasing amount of H_3PO_4 promoted the formation of glycine-valine co-oligomers, whilst the formation of oligoglycines decreased.

On the other hand, the use of H_2SO_4 shows a very different situation (see Fig. 92). Only upon the addition of $2 \cdot 10^{-4}$ moles of H_2SO_4, glycine-valine co-oligomers were formed. These results clearly showed that very different behaviours were obtained depending on the acid used when different amino acids were co-oligomerised. As mentioned before, we were not able to specify the cause for the different behaviours between the two acids employed, and further exploration would be required to understand such differences.

3.3.3 Systematic co-oligomerisation exploration.

In light of the previous results, we decided to systematically explore the formation of peptides between binary combinations of a set of 10 different amino acids in the presence of different acids (Fig. 93). We wanted to investigate whether the co-oligomerisation of amino acids other than glycine was possible, aiming for the formation of peptides incorporating amino acids capable of promoting the formation of secondary structures.

	Gly	Ala	Leu	Asp	Glu	Lys	His	Val	Pro	Thr
Gly										
Ala										
Leu										
Asp										
Glu										
Lys										
His										
Val										
Pro										
Thr										

Figure 93: Co-oligomerisation matrix between 10 different amino acids.

We chose a particular set of amino acids due to their prebiotic relevance: Gly, Ala, Leu, Asp, Glu, Pro, Thr and Val are considered to be prebiotic since they have been synthesised in spark discharge-type experiments and found in comets and meteorites.[227] However, Lys and His were incorporated in these experiments due to their catalytic properties despite not being considered prebiotic amino acids.

All the binary amino acid co-oligomerisation reactions shown in Fig. 94 were performed in the presence of three different acids (HCl, H_3PO_4 and H_2SO_4). In all experiments, amino acids were mixed in equimolar concentrations.

Preliminary HPLC-MS analysis of a co-oligomerisation reaction between Gly and Asp (see Figs. 94&95) revealed a series of peaks that correspond to the mass of G_nD_n oligomers minus the mass of 1, 2 or 3 water molecules. This is likely due to the formation cyclic imide residues (succinimide).[91] Also, since aspartic acid can also react on his carboxylic acid side chain, the formation of β-peptide bonds is also possible (see Fig. 96). Therefore, a particular mass in the spectrum could correspond to oligomers with different ratios of α to β peptide bonds (see Fig.96).

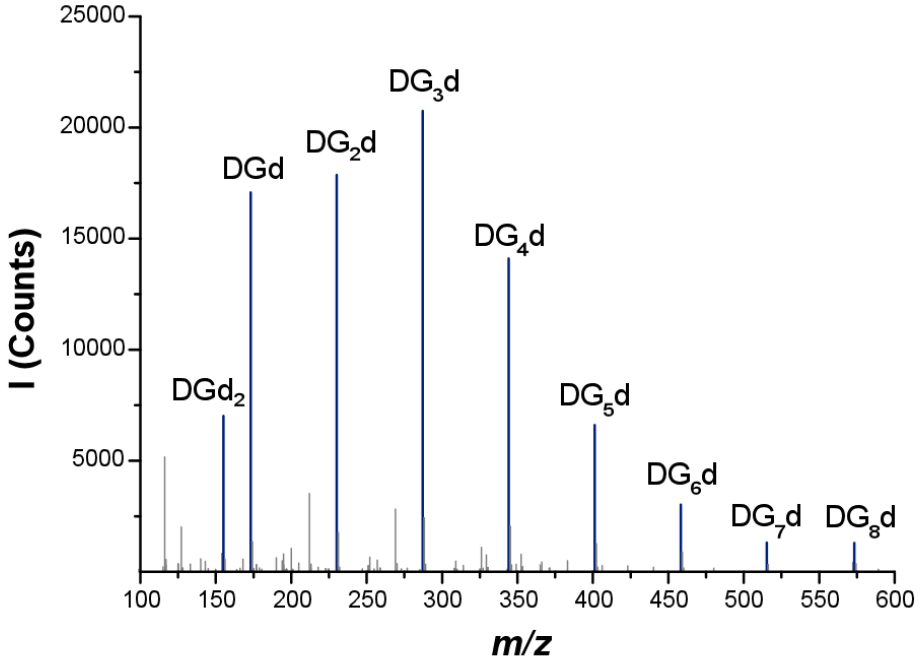


Figure 94: MS spectra of the products obtained in a Gly-Asp co-oligomerisation reaction in the presence of H_3PO_4 corresponding to the integration of the BPC between 2.5 and 5 min. D=Asp, G=Gly, d=dehydration, corresponding to the loss of one H_2O molecule.

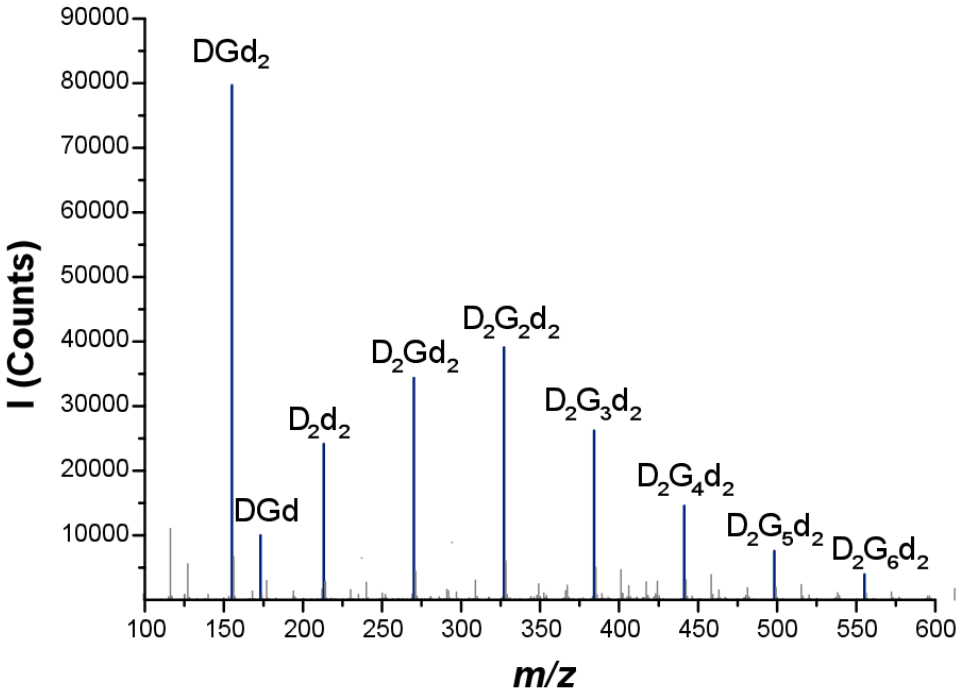


Figure 95: MS spectra of the products obtained in a Gly-Asp co-oligomerisation reaction in the presence of H_3PO_4 corresponding to the integration of the BPC between 7.5 and 9 min. D=Asp, G=Gly, d=dehydration, corresponding to the loss of one H_2O molecule.

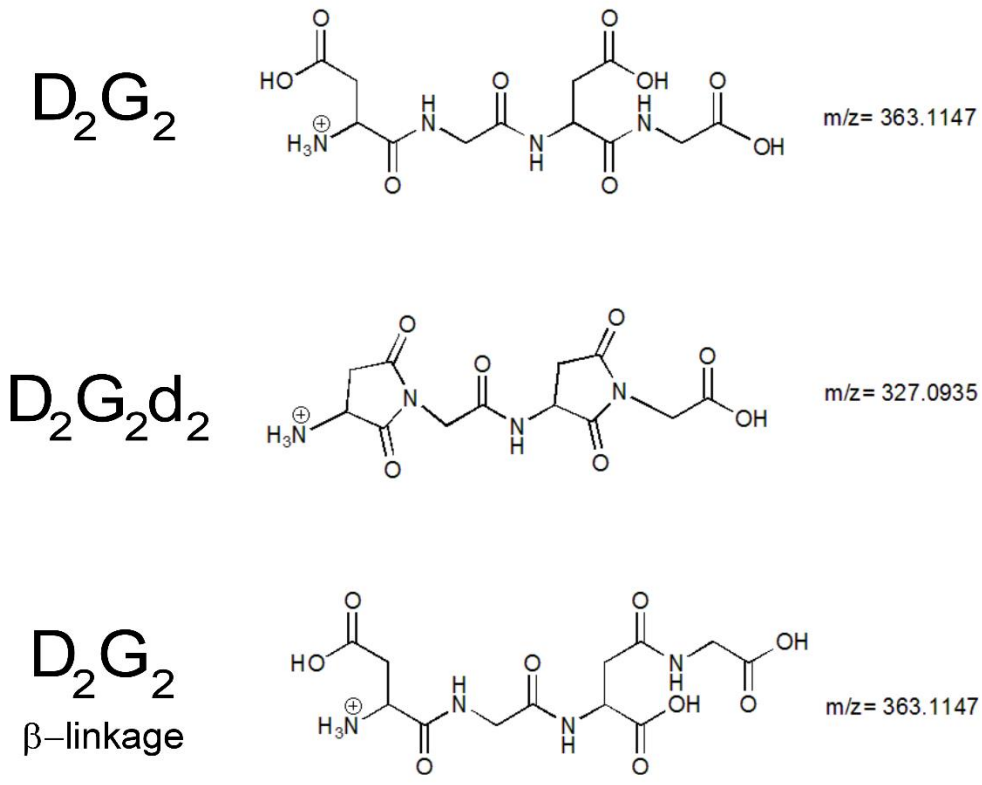


Figure 96: Different examples of structures observed for a tetramer containing 2 Asp and 2 Gly residues. D=Asp, G=Gly, d=dehydration, corresponding to the loss of one H_2O molecule.

The structures represented in Fig. 96 aim to show that many possible different structures can result for a tetramer comprised of 2 Asp and 2 Gly residues (containing cyclic imide residues, β-peptide bonds or a different arrangement of amino acids within the sequence). Whilst the exact nature of the products formed in these reactions was not determined, we note that more than one possible structure could be assigned to a particular m/z value. However, in this set of reactions, the aim was not to determine the exact structure of the oligopeptides formed, but instead, explore the different reactivities between binary combinations of amino acids in the formation of peptide bonds.

Since the products resulting from these reactions were very complex and a large number of data files had to be processed, an analytical workflow was developed by Dr. Andrew Surman to process and represent the products obtained in these reactions.

First, we generated a combinatorial list of all the possible oligomeric products between 3-mers and 15-mers for each co-oligomerisation reaction. Note that peptide products with masses that included the loss of up to 3 water molecules were also considered (see Fig. 97).

Aa composition	**m/z**
GGG-H_2O	172.0714
GGG	190.082
GGV-H_2O	214.1184
GGGG-H_2O	229.0934
GGV	232.129
GGGG	247.104
GVV-H_2O	256.1654
GGGV-H_2O	271.1404
GVV	274.176
GGGGG-H_2O	286.1144
GGGV	289.151
VVV-H_2O	298.2124
GGGGG	304.125
GGVV-H_2O	313.1874
VVV	316.223
GGGGV-H_2O	328.1614
GGVV	331.198
GGVV	343.1364
.	
.	

Figure 97: List of all the possible oligomeric products for a Gly-Val co-oligomerisation reaction. Note that just the first 19 combinations were included for illustration. The total combinatorial list considering just the loss of one H_2O molecule contains 264 possible combinations.

Then, a list of produced oligomers in each co-oligomerisation reaction was generated by comparing the list of combinatorial masses to the oligomeric species detected by HPLC-MS.

First, we wanted to find out the degree of reactivity between different amino acids. Therefore, we represented the combined ion intensities corresponding to peptide species for each co-oligomerisation reaction.

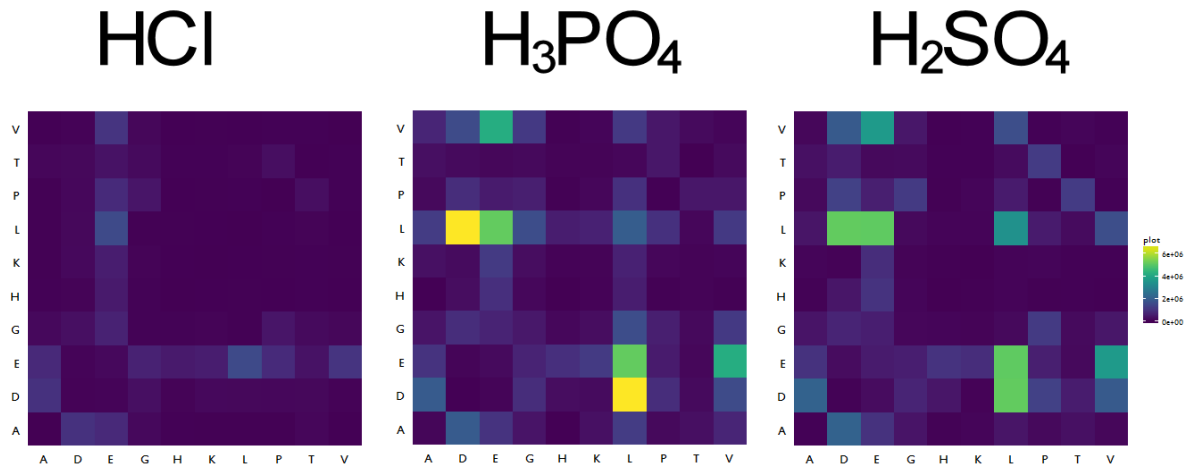


Figure 98: Co-oligomerisation matrix between 10 different amino acids in the presence of three different acids (Colour scale corresponds to combined intensity of product peptide species: yellow=more intense, blue=less intense).

As observed in Fig. 98, different intensities were obtained for particular co-oligomerisation reactions, meaning that specific binary combinations of amino acids reacted more than others to form peptides (The scale bar represents the ion intensities from yellow (more intense) to blue (less intense)). This pattern clearly shows that the formation of peptide bonds from unactivated amino acids is far from a random combinatorial explosion.

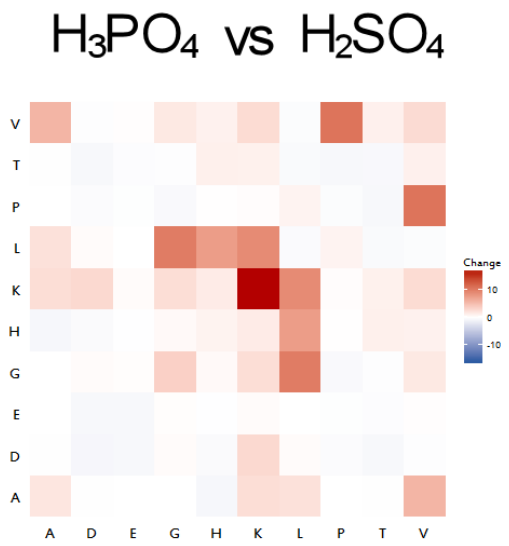


Figure 99: Co-oligomerisation matrix representation of the change in ion intensities between using H_3PO_4 or H_2SO_4 to adjust the pH (colour scale corresponds to degree of change: red=more intense in H_3PO_4 reaction, blue=more intense in H_2SO_4 reaction).

Moreover, upon changing the acid used to adjust the pH, the pattern of reactivity changes. More interestingly, the change in reactivity is different for different co-oligomerisation reactions. Fig. 99 represents the degree of change between the reactions performed using H_3PO_4 or H_2SO_4 to adjust the pH. For example, we can clearly observe that there is a 20-fold increase in the Lys homo-polymerisation reaction by using H_3PO_4. On the other hand, a 5-fold increase is observed in the Asp homo-polymerisation reaction when H_2SO_4 was used instead.

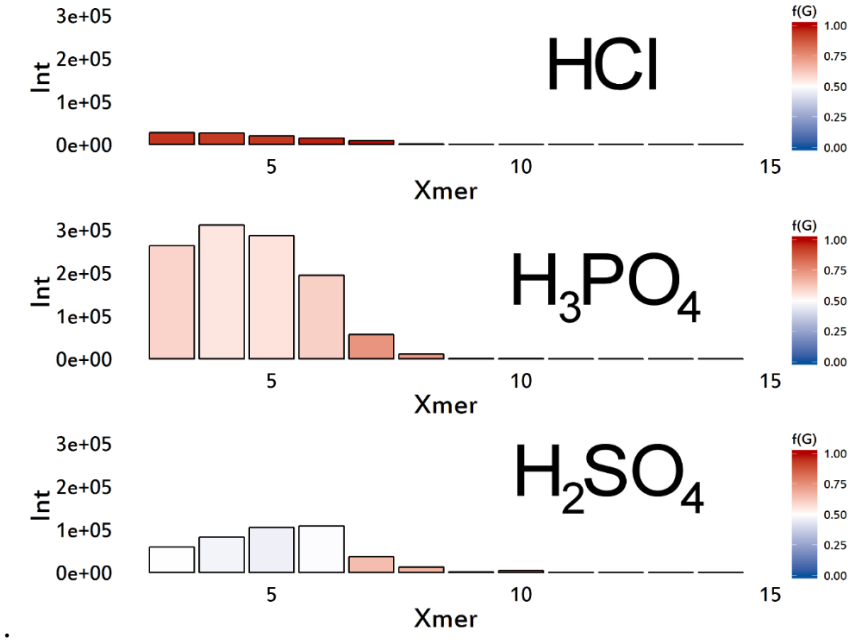


Figure 100: LC-MS BPCs bar graph representation of valine-glycine co-oligomerisation products in the presence of three different acids. The bar on the right represents the identity of amino acids incorporated within the oligomer. (Red=majority of glycine residues, Blue=majority of valine residues).

The different patterns in reactivity represented in Figs.98&99 correspond to the total ion intensities of oligomers formed. However, we also wanted to find out whether the nature of the oligomers produced was also different upon changing the acid used to adjust the pH. As observed in Fig. 100, a dramatic increase in oligomer product formation occurred on changing from HCl to H_3PO_4 in a Gly-Val co-oligomerisation reaction. A moderate increase was observed when using H_2SO_4 instead. More interestingly, the oligomer length distribution was also different: when H_2SO_4 was used, the most abundant oligomers were 6-mers, whilst 4-mers dominated in the H_3PO_4 experiment.

Also, the amino acid composition within the peptide oligomers is also different. For example, whilst the 5-mers obtained when using H_3PO_4 were comprised of a majority of glycine residues, the 5-mers obtained when using H_2SO_4 contained a larger amount of valine residues.

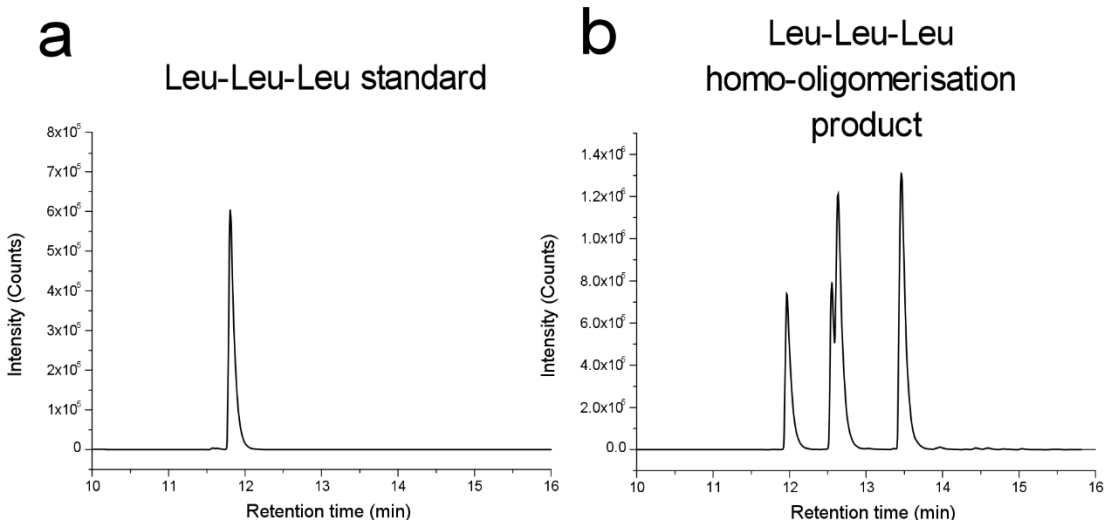


Figure 101: EICs corresponding to Leu-Leu-Leu. a) EICs obtained for Leu-Leu-Leu standard. b) EIC obtained for Leu-Leu-Leu in Leu homo-oligomerisation reaction using H_3PO_4 to adjust the pH.

Further exploration of the EICs in a Leu homo-oligomerisation reaction revealed that for a mass corresponding to m/z= 358.2639, which is assigned to L-Leu_3, more than one species were present. Only one of the four peaks observed in the co-oligomerisation EIC corresponded to L-Leu_3 standard. The same pattern was observed for the other Leu_n oligomers produced in the homo-oligomerisation reaction. This suggests that the other observed peaks in the EIC could correspond to Leu_3 species containing both D and L Leu residues.

Whilst further analysis is required to confirm the presence of D and L amino acid residues, we suspect that racemisation of amino acids at such high temperatures is not unlikely.[228]

In these series of co-oligomerisation experiments, we compared the combined ion intensities that correspond to peptide species and observed different patterns of reactivity when the acid used to adjust the pH was changed. Whilst not an inherently quantitative technique, mass spectrometry allowed us to compare the intensities of peptides products obtained in our reactions. We observed that in the presence of H_3PO_4, the combined ion intensities were greater than the ones obtained when HCl was used. However, we wanted to obtain absolute quantitative information of these reactions, and compare them to the trends observed by mass spectrometry. Since our co-oligomerisation reactions produce a large number of peptide products of similar length and properties, Ion-pairing HPLC could not be employed to separate and quantify the resulting products. Alternatively, we decided to use an HPLC method that measures the unreacted amino-acid monomer instead.

Since amino acids contain amino groups (-NH_2 or -NHR) in their structures, a derivatising reagent that selectively reacts with the amino group was used to quantify the unreacted amount of amino acid monomer (see section 5.2.2.10). In this way, we calculated the monomer to oligomer conversion for each binary amino acid reaction, considering that the amino acid monomers were exclusively reacting to form peptides.

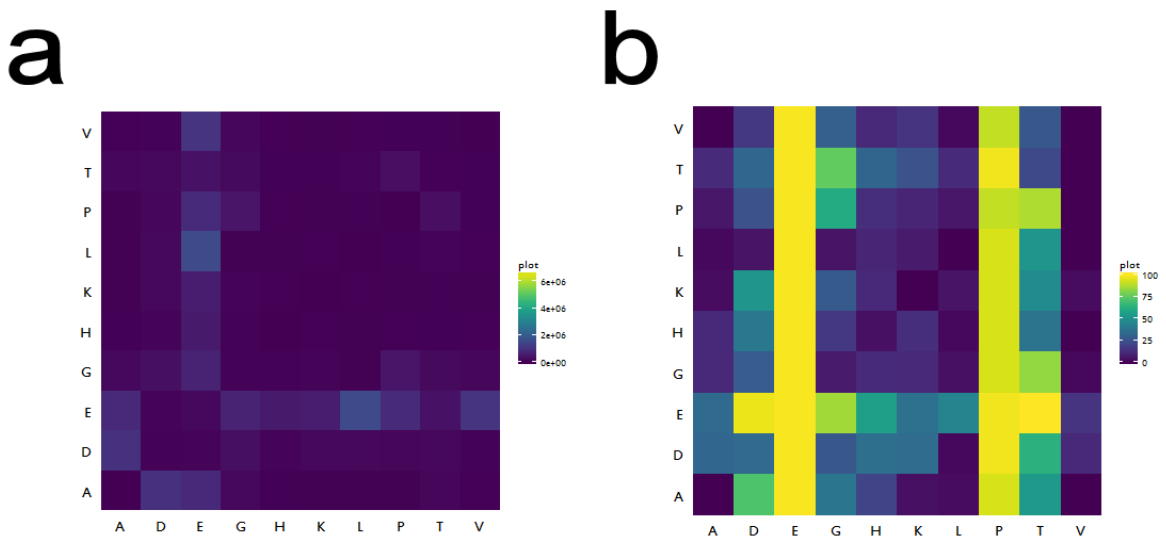


Figure 102: a) Representation of combined ion intensities for peptides species by HPLC-MS analysis of co-oligomerisation reactions. b) Monomer to oligomer conversion values for each binary amino acid co-oligomerisation reaction.

We observed significant differences by comparing the combined ion intensities corresponding to peptides species by HPLC-MS analysis to the monomer to oligomer conversions calculated by HPLC analysis of the unreacted monomer fraction. As represented in Fig. 102 b), glutamic acid and proline show very high conversion upon the reaction with other amino acids. However, in Fig.102 a), we cannot observe high ion intensities corresponding to peptide species. This quantitative analysis has just been performed once without any analytical repeats, and the chromatographic method is still under development to obtain robust and repeatable data. Nonetheless, we note that the formation of insoluble products that are observed for some of the co-oligomerisation reactions might explain the differences between the two analyses. Also, the formation of co-oligomerisation products that do not correspond to peptide species may be possible. Further exploration of the side products that can be obtained in these co-oligomerisation reactions is being carried out.

3.3.4 Environmental programming of peptide synthesis

In the previous section we showed that many amino acids were able to co-oligomerise in the presence of different acids. A thorough mass-spectrometric analysis confirmed that the majority of the products obtained corresponded to peptides species, or peptides containing either branched or cyclic structures. Whilst the conditions used in our experiments are similar to those described in Fox's later work[89], we were able to successfully co-oligomerise various amino acids in the absence of amino acids with electrically charged side chains.

Our next objective was to demonstrate that we could efficiently synthesise oligopeptides of sufficient length such that individual sequence variances could be present in reasonable amounts. Then, we wanted to investigate whether some of these specific sequences have a function (catalytic, structural) that could have conferred an advantage upon an emergent system.

However, the combinatorial explosion that can result from 'one-pot' thermal syntheses of peptides, such as our reaction, has a number of associated problems. For example, if we start an experiment where 6 different amino acids react with 100% efficiency to generate hexapeptides, assuming no sequence selectivity, 6^6 hexapeptides (46656, presuming no epimerisation) would be produced in equal amounts.

If any of the individual hexapeptides had any useful function, they would have to be highly efficient catalysts or extraordinarily tight binders to show any function at such low concentrations. It is clear that reducing the randomness in these reactions is necessary to obtain products in high enough concentration to show any function. But how do we solve this problem?

First of all, in our co-oligomerisation experiments, we demonstrated that the different intrinsic reactivities between amino acids led to a reduced number of possible peptide products. Our next objective was to investigate whether the peptide products obtained in our co-oligomerisation reactions could be 'biased' or 'programmed' by a series of environmental factors, and if different product populations with different functional properties could be obtained upon changes in the environment

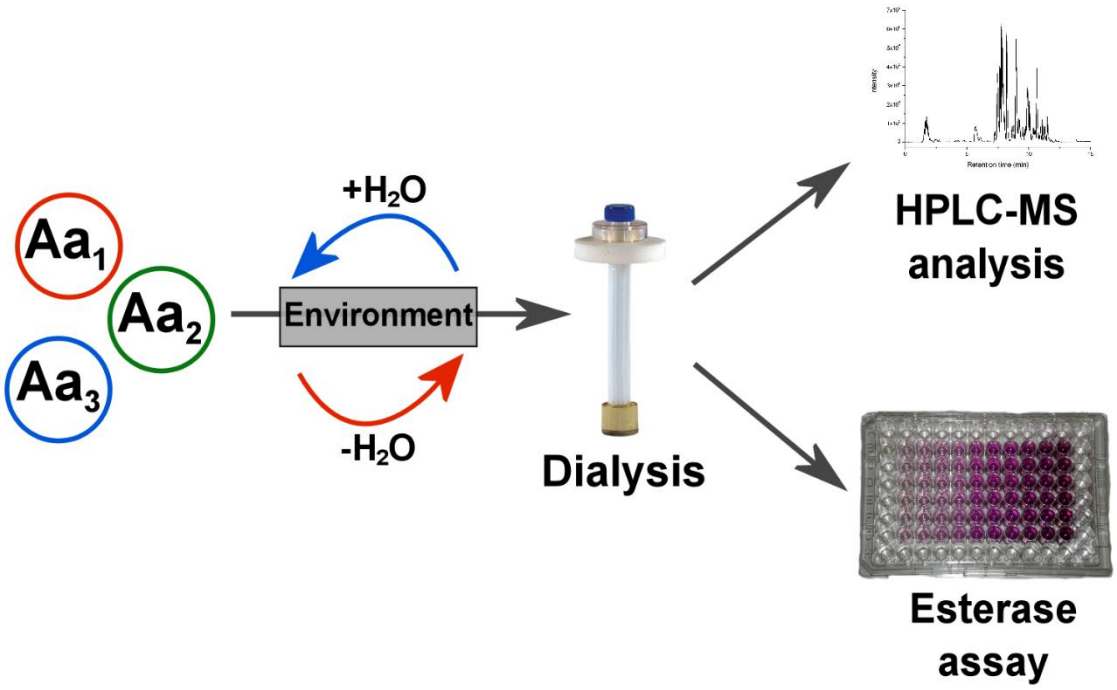


Figure 103: Workflow for the synthesis and analysis of different peptide populations obtained upon changes in the environment.

To do this, we designed a series of co-oligomerisation reactions in which three different amino acids were subjected to different environmental conditions. Then, the obtained products were then filtered and further purified by dialysis to remove the salt/minerals used as 'environment'. The resulting products were then analysed by HPLC-MS. Finally, an esterase activity assay was performed to study whether different product populations had different interactions with the substrate.

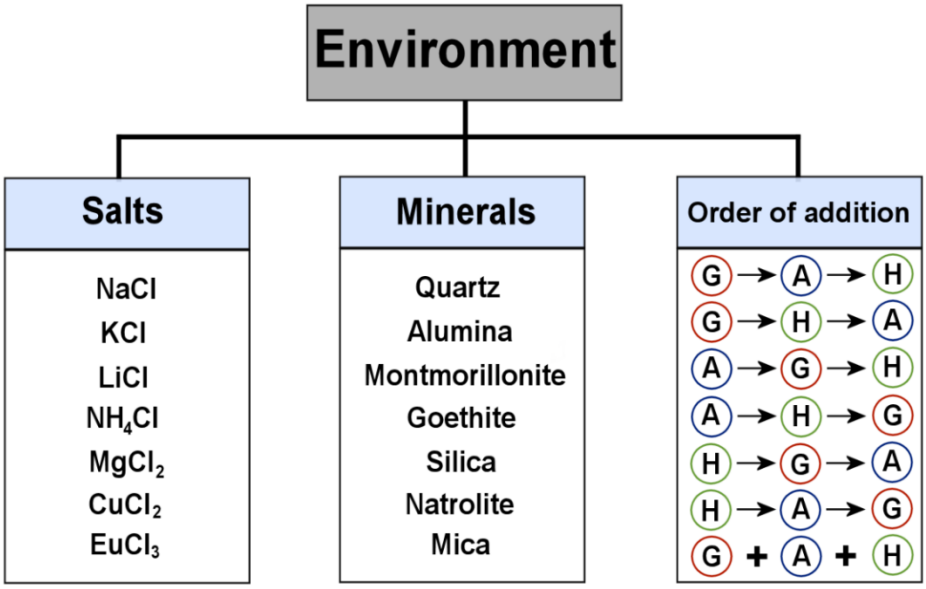


Figure 104: Many different environments (within three different subsets: salts, minerals and order of addition) were chosen to run the co-oligomerisation reaction between three different amino acids (glycine, alanine and histidine).

As shown in Fig. 104, 7 different salts and 7 different minerals were chosen as different environments. A given amount of salt or mineral was added to a vial containing an equimolar mixture of glycine, alanine and histidine. On the other hand, a set of reactions was carried out by adding each one of the amino acids in different orders (note that the final concentration of amino acids was the same at the end of the reaction). In all reactions, a total number of 9 dehydration-hydration cycles were performed. Once the reactions were finished, the obtained products were dialysed using a 100Da cut-off dialysis membrane to separate the unreacted monomers and the soluble salts from the co-oligomerisation products (see section 5.2.5.1 for further details).

The method to dialyse a sample consists in placing the sample inside a dialysis tube, which is then submerged into a beaker filled with deionised water under constant stirring. The water needs to be replaced every 3/4 hours according to the manufacturer, and the total dialysis time is 20 hours. Given the large number of experiments performed, we decided to build a dynamic dialysis set up to speed the dialysis process.

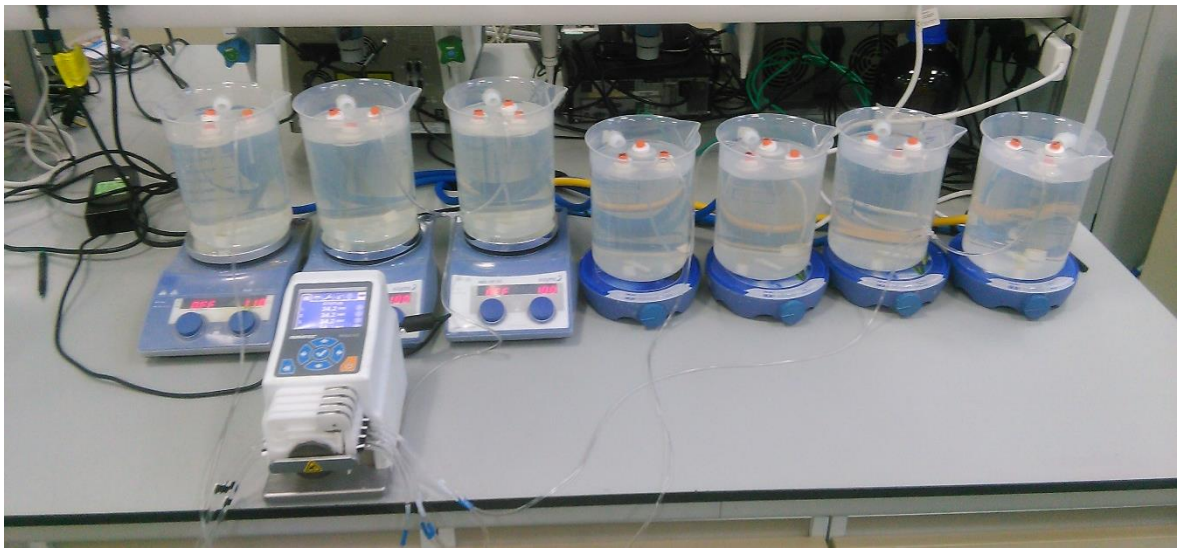


Figure 105: Custom-made dynamic dialysis set up.

As observed in Fig. 105, we connected a peristaltic pump that was constantly flowing deionised water to each beaker at a constant flow rate of 10ml/min. A total of 6l of water was flowed to each individual tank after 20 hours of dialysis.

Once the co-oligomerisation products were dialysed, they were frozen under liquid N_2 and freeze-dried for 48h. Given that the environment could not only change the distribution of the products formed, but also the yield of the reaction, two different approaches were adopted: the resulting dialysed products were either prepared at the same concentration or at the same volume. In the products prepared at the same concentration, a given amount of freeze dried peptide was weighed using a high-precision micro-balance and stock solutions were prepared upon the right addition of water. On the other hand, the products prepared at the same volume were obtained by adding the same amount of water to the total amount of peptide recovered after dialysis. In this way, if the formed products possessed any functional activity, this could be measured as activity/mg of product or on the contrary, activity/reaction.

When the dialysed samples were both prepared at the same concentration and at the same volume, they were then used for HPLC-MS analysis and also for measuring the esterase activity of the dialysed products.

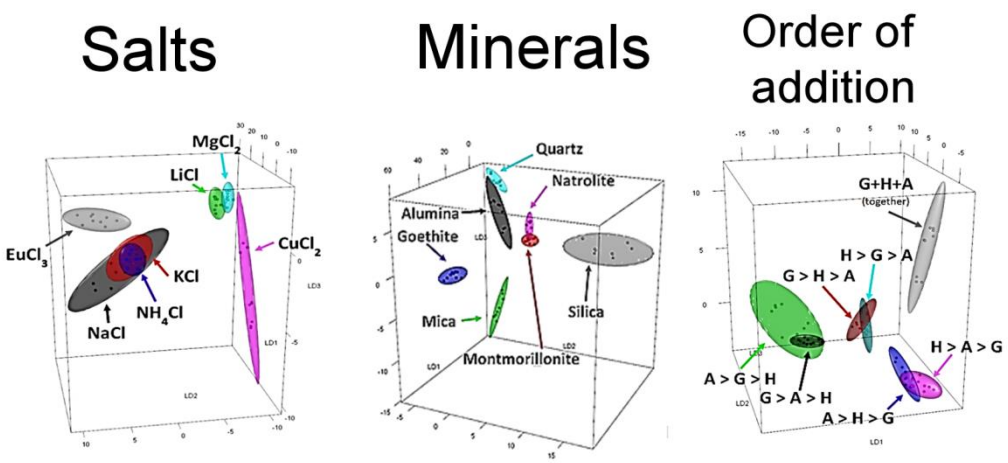


Figure 106: Principal Component Analysis (PCA) corresponding to different peptide populations obtained under different environments.

Principal component and discriminant function (PC-DFA) analysis on the data obtained from HPLC-MS revealed clear differences in peptide product distributions (see section 7.1 for further details). We then wanted to investigate if these differences in product distributions led to different catalytic activities. Therefore, we performed an esterase assay based on the hydrolysis of the ester p-nitrophenyl acetate (pNPA). The hydrolysis of pNPA is easily monitored by UV-Vis spectroscopy due to the formation of the highly coloured product p-nitrophenol (pNP) (see Fig. 107).

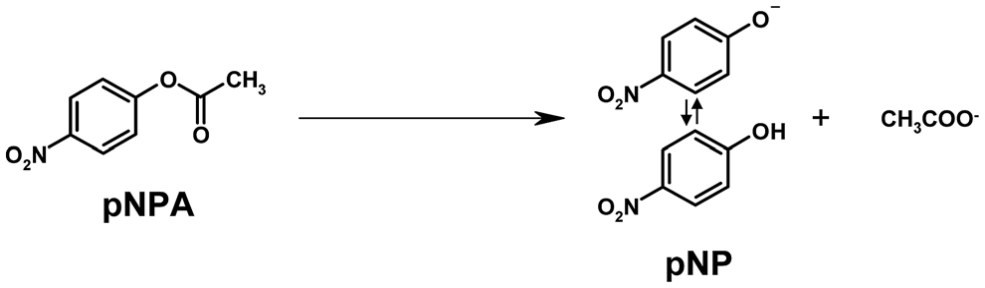


Figure 107: Hydrolysis reaction of pNPA in aqueous solution leading to the formation of pNP and acetate.[229]

p-Nitrophenyl acetate is a very common chromogenic substrate used for determining esterase and lipase activities. Recent work by Rufo *et al.*[225] describe the formation of short peptides that self-assemble to produce highly catalytic amyloids. In this case, they study the hydrolytic activity of very short amyloid peptides (7-mers) by busing p-NPA as a substrate to allow for comparison with natural enzymes. Inspired by this work, we decided to use a similar protocol to determine the different functional properties of the peptides produced in different environments to release p-nitrophenol (see section 5.2.5.1).

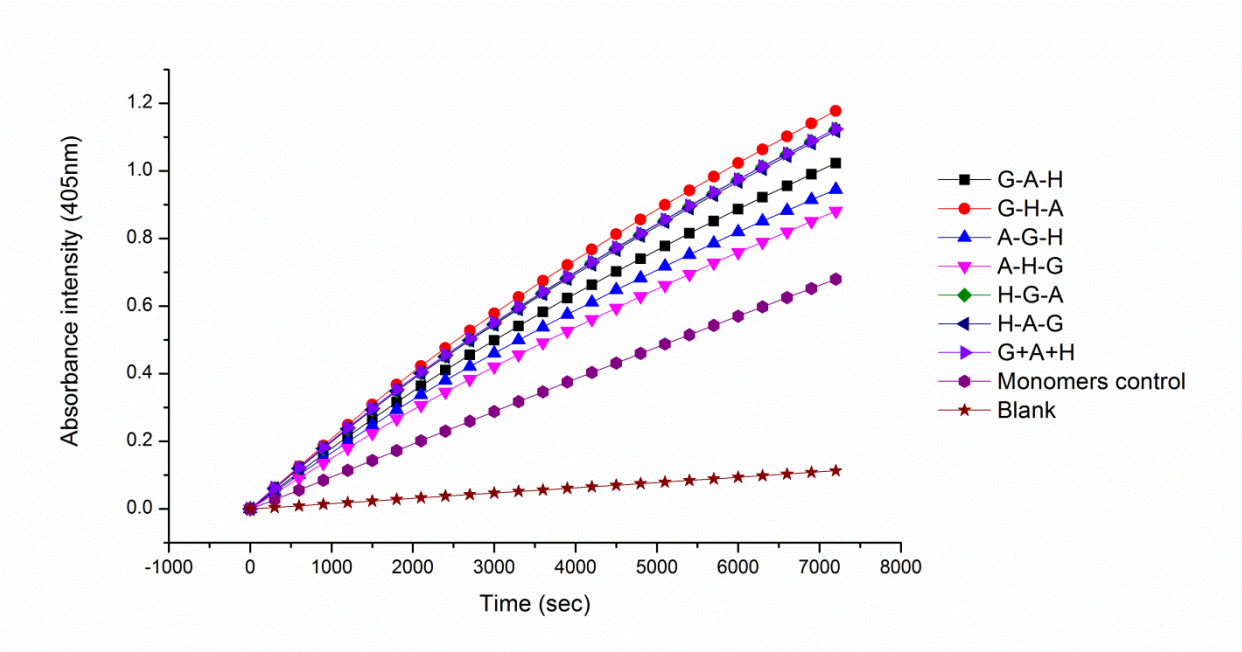


Figure 108: p-NPA assay results for different product populations performed with different orders of addition (same concentration treatment). A control experiment with a mixture of Gly, Ala and His monomers was also performed. See experimental details in section 5.2.5.5.1.

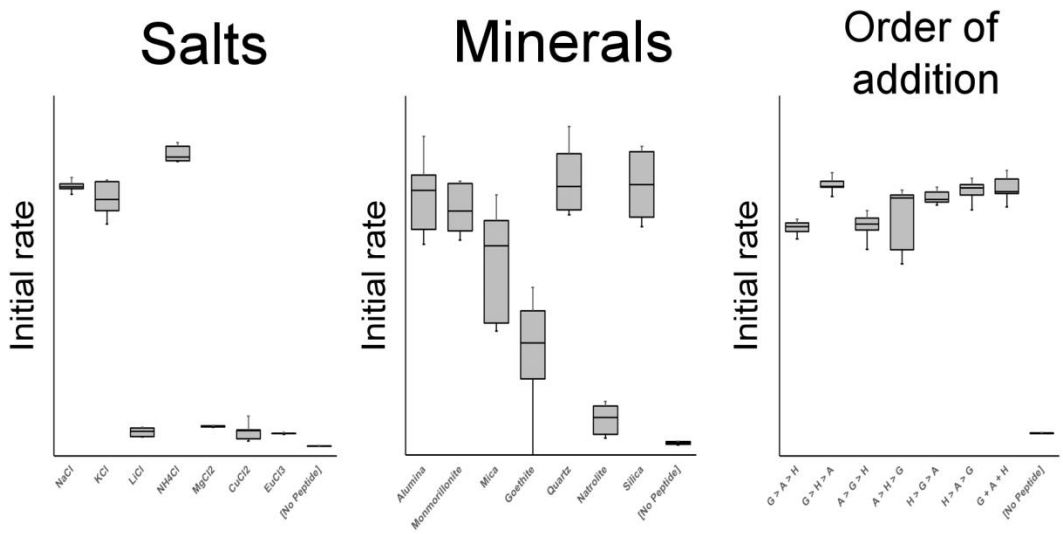


Figure 109: p-NPA assay results for different product populations a) in the presence of different salts (same volume treatment), b) in the presence of different minerals (same volume treatment) and c) performed with different orders of addition (same concentration treatment). Note that the rates represented correspond to the rate of p-Nitrophenol formation for the first 15min of reaction.

When the esterase assay was performed for peptide populations resulting from adding the amino acids in different order of addition, we observed different rates of p-nitrophenol formation between peptide products (see Fig. 108). The esterase assay was then performed in triplicates for each one of the reaction replicates (a total of 9 analyses were performed for each environmental condition).

As observed in Fig. 109, different peptide populations resulting from interactions with different environments interacted with p-NPA to release different amounts of p-nitrophenol. Note that we cannot assure whether this is a catalytic or a stoichiometric reaction. Further investigations, such as calculating the rate of formation of p-nitrophenol at various peptide concentrations will be performed to determine the nature of this interaction.

3.3.5 Section summary

Most of the previous work on attempting to synthesise peptides from unactivated amino acids has resulted on the formation of short peptides in very low yields. Furthermore, the majority of previous research has only studied the co-oligomerisation between a reduced number of amino acids, typically resulting in the formation of dimers and trimers.

Herein, we have systematically explored all the possible binary co-oligomerisation reactions between a set of 10 different amino acids. We have demonstrated that by changing the acid used to adjust the pH, a different pattern of reactivity is obtained. More interestingly, not only does the overall intensity change, but so does the product distribution. However, a thorough investigation of the individual reactivity propensities for each amino acid to react with others would be required to understand these patterns. Preliminary observations show that amino acids with electrically charges side chains co-oligomerise to produce a series of peptides potentially containing cyclic imide residues. We have also observed that amino acids other than Gly homo-oligomerise to produce species that might contain racemised amino acid residues (in the case of Leu).

Moreover, carrying out a co-oligomerisation reaction of three different amino acids in the presence of different environments led to demonstrably different product distributions. Further analysis is being performed to demonstrate that the differences between these peptide populations might not only be in their product distribution, but also in their sequence. Finally, these different product populations also led to different interactions with a chromogenic substrate.

4 Conclusions and future work

Understanding the transition from simple chemicals to self-sustained living systems has been usually attempted from a historical point of view. However, there are still major disagreements regarding the nature of the geochemical environments where life first arose. This has led to the proposal of several models for the Origin of Life, which often lack from experimental support and concordance with environmental constraints.

Over the last decades, traditional synthetic approaches have been used to investigate the formation of life's precursors. Many of these experiments establish their conditions based on the nature of possible prebiotic geochemical scenarios, and are often optimised to make very specific products under very specific conditions.

On the other hand, ahistorical approaches aim to find out the most likely chemical pathways towards the formation of a living system with no constraints on geochemical plausibility. The recent advances in laboratory automation and high-throughput analysis can change the way these investigations are performed, leading to the exploration of a much larger number of possible reactions under different conditions, aiming for simplicity rather than plausibility.

4.1 Microfluidic platform for chemical evolution

Initially, we started to develop a microfluidic platform capable of assisting chemical evolution. This was accomplished by constructing and testing individual microfluidic modules where water-in-oil microdroplets could be selectively generated, fused, split and sorted. This is still an on-going project in which each one of the individual modules will be integrated into a single device (or several interconnected devices) and where encapsulated reagents will be able to become more complex within iterative selection cycles. Our initial aim was to achieve the synthesis of functional polymers from simple building blocks, and then, assist the evolution of these species into more complex molecules with improved functionalities. However, we were not able to find a chemical system that fit the description outlined above. Therefore, we decided to focus our attention on the synthesis of polymers with endowed functionalities that could be later selected by directed evolution within our microfluidic platform.

4.2 Abiotic Peptide Synthesis

Attempting to synthesise peptides from unactivated amino acids under thermal conditions has usually resulted in either the formation of very short oligomers, or in the formation 'proteinoids' that contain a very small percentage of peptide bonds. Since this approach (Fox[35], Rohlfing[86], Meggy[78,79]) failed to produce long peptides, alternative strategies such as activating the amino acids (either using activating reagents or via the formation of Cu complexes) gained more attention for a long time. This was due to the easy polymerisation from N-carboxy anhydrides (by using activating reagents such as CDI[92], EDC[111], or more prebiotically accepted ones like COS[123] or by starting from carbamoyl amino acids(CAAs)[44]). Even when using activating reagents, in the case of COS or CAAs, only very short peptides were obtained.

The most recent literature on peptide formation from unactivated amino acids still faces the same problems encountered 50 years ago. Several examples have reported the formation of very short glycine oligomers (up to 4-mers) when individual conditions were studied: the effect of using salts as dehydrating agents[230], the effect of pH in hydrothermal conditions[105], the effect of adding various minerals[231], etc.

Whilst many of these studies establish their conditions based on the nature of possible prebiotic geochemical scenarios, we decided to take an alternative approach: to explore many different conditions in as many possible combinations with no constraints on plausibility. For that, we constructed the APS (Abiotic Peptide Synthesiser), an automated platform capable of running several experiments in parallel and both vary the chemical inputs and the experimental conditions. By studying the effect of several conditions simultaneously, we were able to synthesise glycine oligomers up to 12-mers in around 50% yield in a single dehydration cycle. We have presented at least an order-of-magnitude increase in yield of higher oligomers compared to previous work.

We first observed that the reaction both proceeded under basic and acidic conditions, and also that by increasing the number of hydration-dehydration cycles we were able to form insoluble products containing glycine oligomers up to 20-mers. We also found that other conditions, such as the rate of evaporation, had a great effect on the reaction yields. Furthermore, upon changing from HCl to H_3PO_4 or H_2SO_4 to adjust the pH, we observed differences in both yield and oligomer length distribution.

This led us to investigate the formation of co-oligomeric products in the presence of these three different acids.

4.3 Programming the formation of heteropeptides

A systematic study of the binary co-oligomerisation reactions between a set of ten different amino acids revealed that peptides are also formed between amino acids that incorporate a wide range of functionality in their side chains: carboxylic acid (Glu and Asp), primary (Lys), secondary (Pro) and aromatic (His) amines, alcohol (Thr) and hydrophobic (Ala, Val) groups. A thorough mass-spectrometric analysis workflow was developed to identify the co-oligomerisation reaction products, which confirmed clear differences in reactivity when the acid used to adjust the pH was changed. A general observation was that by incorporating H_3PO_4 or H_2SO_4, the formation of longer peptides was promoted. Also, not only did the overall intensity change, but so did the product distribution.

The conditions reported in previous experiments (Fox[89] and Varfoloomev[90]), are strikingly similar to our reaction conditions. However, the reason why our approach represents a significant advance over previous work is because we have been able to synthesise long hetero-peptides incorporating various amino acids in the absence of amino acids with electrically charged side chains. Moreover, while previous prebiotic oligomer syntheses have concentrated on bond-forming, here we have concerned ourselves with product distribution and sequence. We demonstrated that ‘uncontrolled’ amino acid condensations show very considerable bias, and that simple treatments can lead to the differentiation of functionally distinct product populations.

4.4 Future work

How could these abiotically synthesised peptides contribute to the metabolic integration of any early entity involved in the chemical evolution process?

Whilst we used a very common chromogenic substrate to study the different hydrolytic activities of different peptide populations, we note that this function might not contribute to the integration of peptides into a chemically evolving system. Further work is now being developed to explore whether some of these peptides have a function (catalytic or structural) that could have conferred an advantage upon an emergent system. For example, we have successfully synthesised hetero-peptides containing aspartic acid and alanine residues (see Fig. 110).

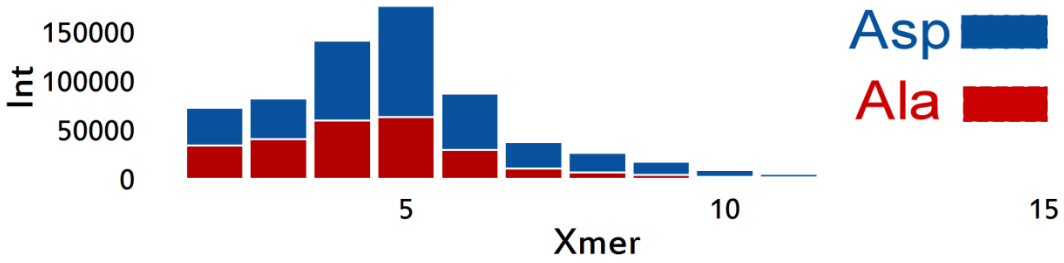


Figure 110: LC-MS BPCs bar graph representation of alanine-aspartic acid co-oligomerisation products in the presence of H_3PO_4.

Previous work by Zhang[232] demonstrated that simple 8-mers with a hydrophilic head composed of aspartic acid and a tail of hydrophobic amino acids such as alanine, valine, or leucine were able to self-assemble to form nanotubes and nanovesicles. These surfactant-like peptides could potentially interact with fatty acid or lipid membranes. For example, a recent study from Szostak[233] described how short amphiphilic peptides were able to localise RNA to the protocell membrane, which could alter RNA reactivity or the strength of RNA-membrane binding.

Alternatively, it has been recently reported that peptides interact with RNA to slow down the reannealing between complementary strands, which then assists several cycles on non-enzymatic template replication. In this publication, Szostak claims that non-coded short peptides could have enhanced early cellular fitness.[234]

It seems clear that abiotically synthesised peptides could have played a key role during the first stages of chemical evolution. Whilst we studied the formation of peptides from pure amino acids, it is more likely that the starting materials in prebiotic syntheses were composed of mixtures of abiotically synthesised precursors. In a recent article, Hud describes the formation of depsipeptides when mixtures of amino acids and α-hydroxyacids were subjected to wet-dry cycles.[235]

Interestingly, a considerable amount of research on the abiotic synthesis of RNA-like oligomers uses very similar conditions to the ones we explored[143–145,236]. Successive hydration-dehydration cycles in the presence of H_3PO_4 or H_2SO_4 have demonstrated to assist the formation of RNA-like oligomers.[237] Therefore, it would be interesting to explore whether the formation of both peptide and RNA oligomers could take place simultaneously. This would be especially relevant if mixed species, like aminoacyl-RNAs or RNA-peptide assemblies, could form and evolve towards the formation of proto-ribosomal machinery.

5 Experimental

5.1 Microfluidics fabrication

5.1.1 Equipment

Bright-field inverted microscope (Olympus).

Spin coater (Laurell, model WS-650MZ-23NPP).

Oxygen plasma cleaner (Harrick-Plasma, Model PDC-001).

UV light source (NewPort, Flood Exposure source, 500 W Hg, model 97435).

Ceramic hot plates.

Oven.

Benchtop centrifuge.

Vacuum desiccator.

LabVIEW software (LabVIEW Core and LabVIEW field-programmable gate array (FPGA); National Instruments).

AutoCAD software (Autodesk).

Silicon wafers (2-inch diameter, Type-P, 1S polished; University Wafer, cat. no. S3P01SP).

Glass slides (TAAB).

Biopsy punches (0.5 and 0.75 mm diameter, Harris Uni-Core)

1/32'' OD PTFE microtubing (Cole Parmer, PTFE#30).

Scalpel.

Plastic Petri dishes.

Wafer-handling tweezers.

Sharp tweezers.

Frosted Scotch tape (3M)

Aluminium foil

5.1.2 Materials

Dry Film Photoresist (DFR) AM-140 and AM-175 (Elga Europe s.r.l., Italy).

MEGAPOSIT® SPR®220 Series Photoresist.

AZ-9260 Photoresists (MicroChemicals GmbH).

SU-8 10 and SU-8 25 Photoresists (MicroChem).

SU-8 developer (MicroChem).

PDMS (Sylgard 184 Dow Corning).

HFE-7500 Oil (3M)

Mineral Oil

EA surfactant (RainDance Technologies).

Low-melting temperature solder: Indalloy 78 Bismuth-Lead-Tin Solder Alloy (Indium Corp.)

Cresol Red

Methylene Blue

Trichloro (1H,1H,2H,2H-perfluorooctyl)silane

1H,1H,2H,2H-Perfluorododecyltrichlorosilane

Acetone

Methanol

Isopropanol

5.1.3 Photo mask design rules

A photomask is an opaque plate or film with transparent areas that allow light to shine through a defined pattern. They are most commonly used in photolithography as templates to transfer patterns to wafers or other substrates in order to fabricate different sorts of devices.

Our photomasks were designed on a vector drawing software. Once the photomasks were designed, they were sent to a photo service company to print them on a high resolution transparency film. First, we used Inkscape vector graphics editor, a very simple and open source vector drawing editor.

However, Inkscape uses the standard SVG (Scalable Vector Graphics) file format, which is not compatible with the file formats that most companies use to process their lithographic data.

We then decided to upgrade our software by purchasing AutoCAD, which is one of the most commonly used programs used for mask design today. The native file format of AutoCAD is DWG, but its interchangeable DXF format is the one required from plotting companies to process and convert the lithographic data into a CAD format called Gerber. Nevertheless, it is unlikely that CAD packages can export as Gerber Format, and for this reason, most companies accept DXF files that they convert into Gerber by using format converters such as LinkCad.

However, some CAD formats contain graphical entities that do not have an equivalent entity in another format. Therefore, there are some rules that should be followed when masks are designed to preserve the photomask data.

- The most important requirement for the conversion is that all entities drawn must be created using closed polygons. There are a few methods to create closed polygons, such as drawing with a zero-width polyline. It is strongly recommended to use the close command to ensure that the figure is closed.

- The creation of composite layers is required for avoiding cross-linking between figures. A composite layer is one that is made up of many other layers with individual properties. Because AutoCAD does not understand the property of a filled shape, closed figures have to be put onto different layers, as each layer in AutoCAD will be translated into a separate file when converted to pattern generation. Otherwise, if all the closed boundaries are on the same layer, the result would be a single filled object.

5.1.4 Dry Film Resist (DFR) Template Development

The following procedure describes the right steps to fabricate a DFR master with an easy and fast technique.

IMPORTANT: The procedure needs to be carried out under red light. Otherwise, the photoresist film can be seriously damaged and affect the quality of the device.
Two different photoresists were used for our DFR development: AM-140 and AM-175 (Elga Europe s.r.l., Italy)

5.1.4.1 Glass slide cleaning

Ten glass slides were set in a glass staining rack and placed in a sonic bath (in DI water/detergent solution) for 20 minutes. Then, the slides were placed again in the sonicator (in DI water only) for 20 minutes.

Finally, the slides were rinsed with acetone, methanol and then isopropanol, and dried with nitrogen gas to ensure complete dryness.

5.1.4.2 Glass slide lamination

A preheated clean glass slide was placed on a hot plate at 125º-150º for 10 minutes, while the laminator was turned on and set up at: 110ºC, speed 2. To ensure a safe lamination, a piece of photoresist film was placed on a small paper sheet. Then, the plastic cover from the matt side of the photoresist film was removed, and a preheated glass slide was placed underneath the film. The slide was carefully laminated and cooled at room temperature for 2 minutes. It is important to ensure that the photoresist film is perfectly adhered to the glass slide. Otherwise, the lamination process must be repeated. The leftover paper section was removed with a scalpel, and the glass slide was heated on a hotplate at 65ºC for 10 minutes. Once finished, the glass slide was cooled at room temperature for 2 minutes.

5.1.4.3 UV exposure

At this point, our photomask template was able to be printed on our laminated glass slide by exposing it to UV. The photomask was placed on the laminated glass slide and exposed to UV light (8W per lamp) for 40 seconds. Once finished, the photo mask was removed and the pattern could be easily observed printed on the photoresist film. It is highly recommended to soft bake the laminated glass slide before exposure at 65° for 10 minutes.

5.1.4.4 DFR development

This is one of the most sensitive steps of the procedure. Care must be taken during sonication to prevent the template being removed from the slide. The developer solution used was 0.8-1 % Na_2CO_3. A big petri dish was filled with developer solution and placed in the sonicator. A previously exposed glass slide was then treated by the following DFR development procedure:

i) Sonicate face down for 60s.
ii) Sonicate face up for 50 s.
iii) Leave the glass slide in solution for 45 s without sonication.
iv) If the slide is not completely developed, we start another cycle by following the instructions

$$t=t_0\text{-}n*10s \text{ where } n=\text{number of cycles}$$

Once the slide was completely developed, both sides of the slide were carefully rinsed with DI water and dried with compressed air. The developed slides were placed in a desiccator chamber together with a drop (or two) of Trichloro(1H,1H,2H,2H-perfluorooctyl)silane on a cover glass. To ensure a complete silanisation, the slides needed to be under vacuum until the silane drop evaporated (approx. 20 minutes). Finally, the vacuum was turned off and left on for another 20-30 minutes. Developed slides were finally hard baked on a hotplate at 120 °C for 2-18 hours

5.1.5 SU-8 10 fabrication process

The following procedure goes through the 6 steps required for the fabrication of any photoresist master. In this case, a detailed protocol for the fabrication of a 20µm thick SU-8 10 is described.

5.1.5.1 Wafer cleaning

Although Si wafers must arrive clean, sometimes they can contain some traces of organic material on the surface, which can lead to some undesired interaction with the photoresist. To avoid this, wafers were rinsed thoroughly with acetone, methanol and isopropanol. Immediately after, the isopropanol was blown with pressurized nitrogen. Finally, in order to remove any leftover solvent, the wafer was placed on a contact hotplate at 200 °C for at least 5 minutes.

5.1.5.2 Spin process

In first place, the bowl of the spin-coater was entirely covered with aluminium foil. Then, a silicon wafer was placed and aligned into the vacuum adapter. Once the desired spin parameters were set up, approx. 3ml of SU-8 10 were poured onto the wafer. It is highly recommendable to pour carefully the photoresist in order to avoid the formation of air bubbles on the surface. Finally, the vacuum was engaged and the resist was spun in two cycles: 1) 500 rpm for 15s to spread the resist and 2) 40 seconds at 1600 rpm to achieve the desired thickness.

5.1.5.3 Pre-bake and Soft-Bake

The wafer was pre-baked at 65 °C for 2 minutes, starting from 40 °C and ramping the temperature 5 °C/min, followed by a soft-bake of 2.5 minutes once reached 95° C.

It is important to let the wafer cool to room temperature before exposure (10 min approx.) Cracks in the SU-8 film may appear if the wafer is cooled down rapidly.

5.1.5.4 Exposure

Firstly, the UV source was turned on 15 minutes prior exposure for stabilisation. The right exposure parameters were set up: 3s at 300W exposure dose.

The emulsion side of the film photo mask was carefully placed onto the wafer, trying to avoid any dust particle on the surface.

Then, a 4 inch quartz disc was placed on top of the photo mask and the UV source was turned on. UV protective eyewear must be worn during this step.

5.1.5.5 Post-exposure bake

The wafer was pre-baked at 65 °C for 1 minute, starting from 40 °C and ramping the temperature 5 °C/min, followed by a soft-bake of 2.5 minutes at 95° C. Immediately, the wafer was let to cool down to room temperature on a flat surface.

5.1.5.6 Wafer development

Once cooled down to room temperature, the wafer was transferred to a crystallising dish filled with PGMEA (SU-8 developer). Once the wafer was completely submerged, the dish was constantly agitated for ~2 min until the appearance of the desired features. When the undesired photoresist was removed, the wafer was rinsed with isopropanol. The appearance of a whitish residue on the surface of the substrate indicates underdevelopment; therefore, the wafer needs to be submerged into the PGMEA solution until proper development. Finally, the wafer was carefully dried with pressurized nitrogen.

5.1.5.7 Wafer silanisation

The wafer was placed in a desiccator with a drop (or two) of Trichloro(1H,1H,2H,2H-perfluorooctyl)silane on a cover glass. To ensure a complete silanisation, the desiccator was left under vacuum for 15min. Then, the vacuum was turned off but the wafer was left in the desiccator for 30 more minutes.

5.1.6 SPR220-7.0 fabrication process

Same steps were performed as the ones described in the previous section. However, some of the conditions changed, as summarized in this table:

SPR220-7.0 fabrication process for a 40µm thick layer		
Spin Coat	1000rpm for 40 seconds	Spin coat twice with a soft-bake step in-between
Pre-bake	95 °C for 100 seconds	Ramping the temperature 5 °C/min
Soft-bake	95 °C for 115 seconds	Ramping the temperature 5 °C/min
Re-hydration	24 hours hold time	
Exposure	250W for 38 seconds	Let the wafer sit 30 minutes before development
Development	Agitate for 5-10 minutes	

Table 1: SPR220-7.0 fabrication process for a 10µm thick layer.

5.1.7 AZ-9260 fabrication process

Spin Coat	2400 rpm for 60 seconds	
Soft-bake	110 °C for 165 seconds	Ramping the temperature 5 °C/min
Re-hydration	30 minutes hold time	
Exposure	250W for 38 seconds	Let the wafer sit 10 minutes before development
Development	Agitate for 2-3 minutes	

Table 2: AZ-9260 fabrication process for a 10µm thick layer.

5.1.8 PDMS fabrication

1. PDMS silicone elastomer (Silicone resin solution, PDMS Sylgard 184 Dow Corning.) and curing agent (Dimethyl, methylhydrogen siloxane) were mixed in a 10:1 ratio for 5 minutes (20g of PDMS per 2g of curing agent). Curing agent was firstly weighed, and then the polymer base was added.

2. The pre-cured PDMS mixture was stirred with a spatula for 5 minutes. It was needed to both swirl and fold the mixture to ensure that the curing agent was evenly distributed.

3. The mixed pre-cured PDMS was then degassed by placing it in the vacuum desiccator and evacuating the chamber. When bubbles appeared, they rose to the surface of the mixture and popped. The mixture was degassed for a minimum of 30 min, and vented when bubbles come close to the surface.
   Alternatively, PDMS could be centrifuged at 4.4 krpm for five minutes to degas the mixture.

4. PDMS mixture was poured onto our developed and previously silanised silicon wafer, being careful not to trap any air bubble. (If necessary, mixture was degassed another time), and baked on a hotplate at 65 °C for 2 hours (until it solidifies).

5. At the end, the device was cut with a scalpel, and carefully, peeled from the template.

6. When PDMS was peeled off from the template, the released layer was placed on a clean surface with channels facing upwards. The tip of the biopsy punch was wiped off with ethanol and brought into alignment with the port. We usually used 0.75mm OD biopsy punches. We needed to adjust the puncher so that it was as vertical as possible. The punch was pushed through the PDMS until hitting the bottom. The punched section from the underside of the device was discarded using a pair of forceps. This procedure was repeated for each input. Finally, the punched PDMS device was placed on a clean surface with features side up, and it was ready to bond to a clean glass slide.

7. Both a clean glass slide and the PDMS device were exposed to plasma treatment for 45 to 60 seconds. Once finished, both surfaces were brought together, and slightly pressed against each other to remove any trapped air bubbles. A post-exposure bake at 65 °C for 30 min is recommended to promote a stronger bonding.

5.1.9 PDMS silanisation

Just after PDMS to glass bonding, a 1% (vol/vol) perfluorododecyltrichlorosilane solution was flushed through the microfluidic channels using a 1ml syringe with a 0.22µm PTFE filte. The solution was left to react inside the microchannels for 30 seconds. Pressurised nitrogen was then flushed through each one of the inlet ports, and neat HFE fluorinated oil was then used to rinse the microchannels to remove the unreacted silanes. The silanised microfluidic devices were then sealed for storage by placing a piece of scotch tape on each one of the inlets.

5.1.10 Multilayer Soft Lithography fabrication

Two different photoresist masters are required for the fabrication of a multilayer device: a flow master and a control master. The flow master consisted on a positive photoresist patterned structure (AZ-9260), whereas the control master was fabricated by using a negative photoresist (SU-8 10).

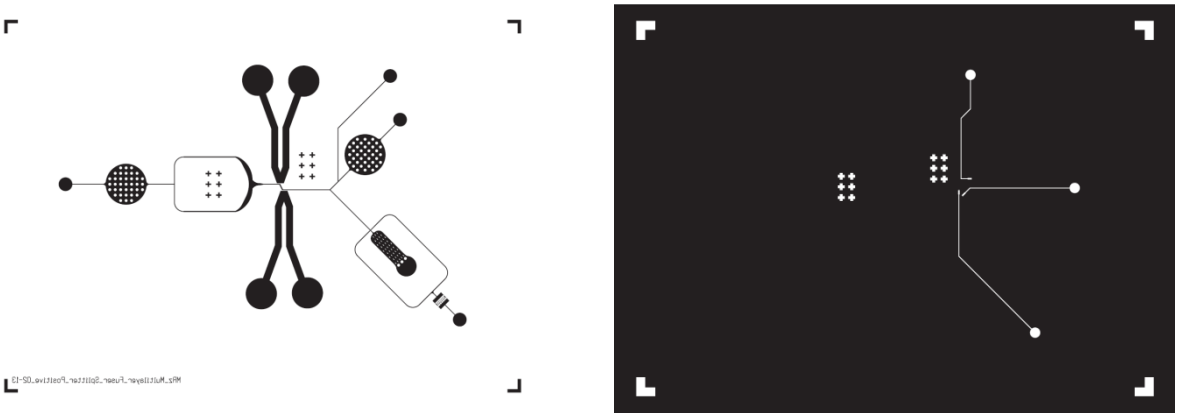


Figure 111: Left) Flow-layer Positive Photomask CAD design. Right) Control-layer Negative Photomask CAD design.

Two main approaches have been described for the fabrication of monolithic micromechanical valves in a PDMS chip: Push-up and Push-down valves.

In our initial experiments, we fabricated Push-up valves by bonding the flow layer on top of the control layer.

For the fabrication of the flow layer, ~15g of a 10:1 PDMS mixture were poured onto the flow master. The mixture was then cured at ~65 °C for two hours, and the PDMS slab was peeled off from the master. Inlet holes were drilled by using a 0.75mm biopsy punch.

Alternatively, the control layer was fabricated by spin-coating 1g of a 1:10 PDMS mixture at 1000 rpm for 5 minutes to achieve a 20 µm thick layer. The wafer was placed on a hotplate at 65 °C for one hour until the PDMS was cured.

Both PDMS layers were oxygen-plasma treated for 45 seconds. Prior to bonding them, three or four drops of water were spread onto of the control layer. Then, the flow layer was placed on top of the control layer carefully, trying not to trap any air bubble between the two layers.

An optical microscope was used to align the flow layer onto the control layer.

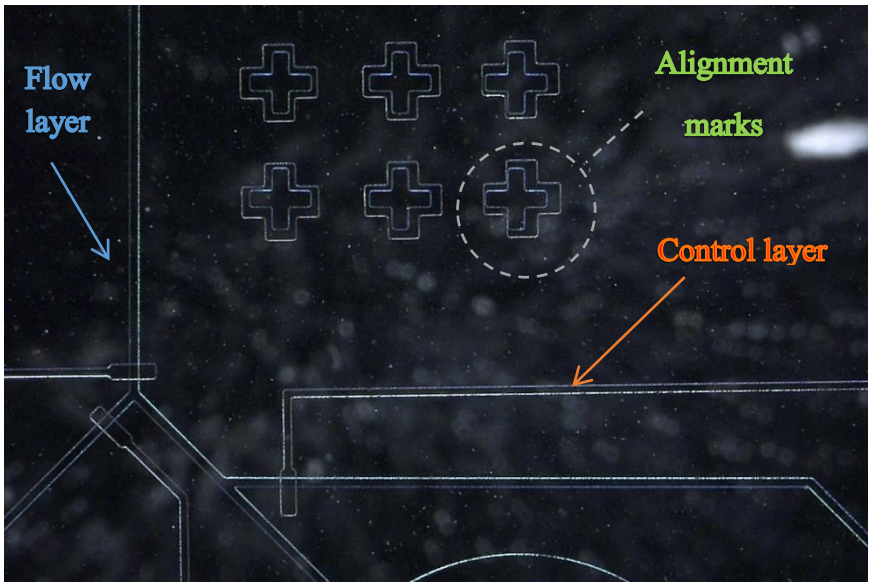


Figure 112: This photograph shows the alignment of a PDMS flow layer on top of a control layer. Although being plasma –treated, PDMS layers were not irreversibly bonded due to the presence of a thin water layer in between, so the top layer could slide on to the control layer until achieving the right positioning of the micro valves.

The aligned layers were placed in a desiccator for 30 minutes to remove the remaining water layer. Immediately after, bonded layers were placed in an oven at 75 °C (1 hour) and 100 °C (1 hour) to ensure proper bonding.

Finally, the PDMS device was peeled off from the control master, and the corresponding holes for the inlet valves were drilled using a 0.75mm biopsy punch. The device was then bonded to a clean glass slide as described in section 5.1.8.

Occasional collapsing of the flow channels occurred when both layers were put together. To avoid this, PDMS mixtures were alternatively cured at 100 °C to harden the elastomer.

5.1.11 Electrode fabrication

A microfluidic device was placed on a hotplate at 90 °C for 5 min. Then, a 5cm piece of a low-melting temperature solder wire was cut and inserted into the electrode microchannel. Upon contact with the glass slide, the wire melted and filled the entire electrode microchannel. The wire was pushed until the melted solder came out from outlet. Note that it is common to trap air bubbles during this process. Any trapped air bubbles would prevent a proper electrical contact. To avoid this, the microfluidic device was previously placed in the desiccator for at least 30 minutes. Finally, metal pins were quickly inserted into each one of the electrode inlet port when the solder was still melted. Upon cooling, electrical contact was verified with a multimeter.

5.1.12 Single-layer valve and Ensemble Sorter device fabrication

The challenge in the fabrication of high-aspect ratio structures required to produce sing-layer micromechanical valve resides in achieving a tick enough-layer of photoresist that can be developed without complications. In our first attempts, we decided to use SU-8 25 and spin it at the lowest speed (750rpm) to achieve a 50µm photoresist thick layer. In all cases, the PDMS used was 1:20 (curing agent : elastomer silicone).

Spin Coat	750 rpm for 40 seconds	
Soft-bake	65 °C for 5 min + 95 °C for 15 min	Ramping the temperature 5 °C/min
Exposure	300W for 10 seconds	Let the wafer sit 10 minutes before development
Post-Exposure bake	65 °C for 1 min + 95 °C for 5 min	
Development	Agitate for 2-3 minutes	

Table 3: SU 8-25 fabrication process for a 50µm thick layer

These conditions were good enough to generate a photoresist master capable of producing 20µm thin PDMS membranes with enough high aspect ratio to create a functional micromechanical valve. However, in the case of the Ensemble Sorter, the dimensions of the channels were not appropriate to efficiently sort highly-packed populations of droplets. Therefore, a new protocol was developed to fabricate a photoresist master with a higher channel aspect ratio. In this case, a photoresist supplied by Tokyo Ohka Kogyo Europe B.V., called TMMR was capable of achieving the formation of up to 120 µm thick layers in a single-step fabrication process. A fabrication protocol was optimised to achieve the formation of a 80µm thick photoresist layer.

Spin Coat	500 rpm for 40 seconds	
Soft-bake	60 °C for 10 min + 100 °C for 20 min	Ramping the temperature 5 °C/min
Exposure	300W for 10 seconds	Let the wafer sit 10 minutes before development
Post-Exposure bake	95 °C for 10 min	
Development	Agitate for 20 minutes	

Table 4: TMMR fabrication process for a 80µm thick layer.

5.2 Abiotic Peptide Synthesis

5.2.1 Materials

All solvents used were HPLC grade, 99.9% (VWR). Glycine, L-Alanine, L-Leucine, L-Aspartic Acid, L-Glutamic Acid, L-Lysine, L-Histidine, L-Valine, L-Proline and L-Threonine were purchased from Sigma Aldrich.

Sodium tetraborate decahydrate (>99.5%), Sodium phosphate dibasic (>98.5%) and Sodium azide (>99.5%) were purchased from Sigma Aldrich. OPA Reagent, FMOC Reagent, Borate buffer 0.4N pH=10 were purchased from Agilent Technologies. HEPES 1M buffer was purchased from ThermoFischer Scientific. p-Nitrophenyl acetate was purchased from Sigma Aldrich.

5.2.2 Instrumentation

5.2.2.1 IP-HPLC analysis

Ion-pairing High Performance Liquid Chromatography analysis was performed using an Agilent 1100 HPLC system fitted with a reversed phase C18 column (Phenomenex Luna, 300 x 7.8mm). Samples were injected in 5µL aliquots, and eluted isocratically at 0.3 ml/min with a mobile phase consisting of 50mM KH_2PO_4 and 7.2mM $C_6H_{13}SO_3Na$ solution, adjusted to pH 2.5 using H_3PO_4. The oligomeric products were detected at 195 nm and the retention times confirmed by comparison with pre-made standards containing glycine monomer, glycine anhydride as well as glycine peptide oligomers (n = 2-6). The instrument was controlled, and the resulting data analysed using Agilent Technologies OpenLAB Software.

5.2.2.2 Size-exclusion Mass Spectrometry (SEC-MS) analysis

SEC-MS analyses were performed (also with an Agilent 1100) system fitted with a Polysep-GFC-P 1000 column (Phenomenex, 300 x 7.8mm) and PolySep-GFC-P guard column (Phenomenex, 35 x 7.8mm). Samples were injected in 10µL aliquots, and eluted isocratically at 0.4ml/min with 90% aqueous NH_4OAc (50mM) and 10% acetonitrile.

The MS apparatus was a Bruker MaXis Impact instrument, calibrated for the 50 – 1200Da range using sodium formate solution. The eluent stream was introduced directly into the source (no splitting) following the DAD detector, at a dry gas temperature of 200 °C. The ion polarity for all MS scans recorded was positive, with the voltage of the capillary tip set at 4800V, end plate offset at −500V, funnel 1 RF at 400Vpp and funnel 2 RF at 400Vpp, hexapole RF at 100Vpp, ion energy 5.0 eV, collision energy at 5 eV, collision cell RF at 200Vpp, transfer time at 100.0µs, and the pre-pulse storage time at 1.0µs.

5.2.2.3 RP-HPLC-MS

Reverse phase High Performance Liquid Chromatography analyses were performed (also with an Agilent 1200 series instrument) fitted with an Agilent Poroshell 120 EC-C18 (4.6 x 50mm, 2.7µm) column. Samples were injected in 10 µL aliquots (except when required to obtain higher signal intensity for some MS/MS experiments – all chromatograms obtained with comparable injection volumes) and eluted with a linear gradient mixture of solvents A (water w/0.1% *v/v* formic acid) and B (100% acetonitrile w/0.1% *v/v* formic acid) over 21 min as follows: 0min – 100% A; 3min – 100% A; 13min – 100% B; 15min – 100% B; 18min – 100% A. The column over was maintained at 30 °C.

The MS apparatus was a Bruker MaXis Impact instrument, calibrated for the 50 – 1200Da range using sodium formate solution. The eluent stream was introduced directly into the source (no splitting) following the DAD detector, at a dry gas temperature of 200 °C. The ion polarity for all MS scans recorded was positive, with the voltage of the capillary tip set at 4800V, end plate offset at −500V, funnel 1 RF at 400Vpp and funnel 2 RF at 400Vpp, hexapole RF at 100Vpp, ion energy 5.0eV, collision energy at 5eV, collision cell RF at 200Vpp, transfer time at 100.0µs, and the pre-pulse storage time at 1.0µs. In MS/MS experiments CID energies were optimised according to products (typically between 15 and 30eV), and the quadrupole mass ‘window’ was set as 0.01Da.

5.2.2.4 General mass spectrometry control and data processing

All data acquired using the Bruker MaXis Impact instrument, controlled by the Compass software suite; where chromatography analysis was added, this process was controlled by accompanying Bruker Hystar software (running Agilent ICF for instrument interface). Where extracted ion chromatograms are presented, they were extracted from the raw data using Compass Data Analysis (m/z +/- 0.01, unless otherwise stated) and exported as .xy data; they were then plotted using Origin 8.5.0 (applying Savitsk-Golay smoothing – points of window = 5, polynomial order = 2; normalising intensity to a 0 – 1 scale for ready comparison). MS/MS spectra presented in this thesis were exported from Compass Data Analysis plotted in R, and theoretically predicted fragmentation patterns (based on putative sequence assignment) were compared with experimentally-derived spectra using the OrgMassSpecR library (*m*/*z* matching +/- 0.05; threshold intensity 3%) running in the R environment.[238] MS/MS spectra presented in the manuscript were exported from Compass Data Analysis, and underwent the same assignment process, however were plotted for publication using Origin 8.5.0.

5.2.2.5 Generation of vBPC traces to demonstrate co-oligomerisation

Observed base peak chromatogram (BPC) data for RP-HPLC-MS runs were produced to demonstrate co-oligomerisation of amino acid species to form complex mixtures. These were constructed from a combination of all the observed extracted ion chromatograms (EIC) corresponding to the members of a list of possible oligomer products. This was accomplished in the following steps (all operations in R using a custom script): a) Acquisition of RP-HPLC-MS data for a set of amino acid oligomerisaton products. b) Conversion of the raw data files from Bruker proprietary format to centroided .mzXML format, using the MSConvert software.[239] c) Generation of a combinatorial list of all the putative oligomeric products of the condensation reaction (ie. all combinations and sequence permutations) under R (utilising the OrgMassSpecR library for fragmentation pattern prediction and mass calculation), produced on the basis of the amino acids incorporated and a maximum length (chosen to ensure that all species up to the mass limit at which the data was acquired was accounted for). d) Removal of homo-oligomer products (i.e. all VBPCs represent only co-oligomerisation of the starting species). e) Generation of the putative of a mass list incorporating all the putative species.

f) Extraction of EIC data for each member of the mass list using xcms[240] under R. g) Generation of a vBPC from the EICs based on the highest intensity observed in the EIC of any member of the mass list and h) Plotting in R.

5.2.2.6 MALDI analysis of a glycine homo-oligomers solid fraction

Matrix Assisted Laser Desorption/Ionization (MALDI) analyses were performed using a Voyager DE-STR MALDI-TOF with Linear and Reflectron analysers (equipped with an Nd:YAG laser (λ = 355nm)). Sample was attempted by solvent-free MALDI, initially with CHCA (α-Cyano-4-hydroxycinnamic acid) matrix typically used for peptide samples. However, no polymeric species and only matrix ions were observed. The analysis was repeated with DHB (2,5-dihydroxybenzoic acid) matrix, and now in addition to matrix ions, 4 oligomeric series were observed. The four series all have the expected repeat unit mass of 57Da, and correspond to $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, and the MALDI artefact $[M-H+2Na]^+$; species for n=8 can be observed considering only–H and –OH endgroups. Despite these four series splitting the ion intensity, oligomeric species are observable up to around 1300Da, corresponding to at least $(Gly)_{20}$.

5.2.2.7 Transmission IR Spectroscopy

The solid samples collected for IR analysis were obtained by the following experimental procedure:

Glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was diluted to a total final volume of 4 ml and the pH adjusted to 9.5 using NaOH. Two different experiments were simultaneously performed with (1ml of a 1M NaCl solution) and without NaCl. Both samples were left to dehydrate at 130 °C for 15h, in order to evaporate the solution to complete dryness. Once a cycle was finished, the process was re-started by re-hydrating the sample with 4 mL of HPLC water, for a total number of 10 dehydration-hydration cycles. Insoluble oligomeric products were separated from the supernatant by centrifugation and decantation. The solid was further washed by re-suspension in HPLC grade water and freeze-dried overnight. Transmission IR Spectroscopy was performed on samples in the solid phase using a Thermo Scientific Nicolet iS5 instrument with Specac Golden Gate attachment and processed with OMNIC software.

5.2.2.8 ^{1}H-NMR Spectrometry

^{1}H-NMR spectra were recorded on a Bruker Avance III (500.2 MHz) using the deuterated solvent as the lock and the residual solvent as the reference. All spectra were run in D_2O. For NMR analysis, deuterated DMSO was added to a sample from one of the peptide formation experiments and the mixture heated to aid dissolution. The saturated DMSO was then filtered into an NMR tube and a proton spectrum was taken. Using similar methods, saturated standards of the glycine monomer, dimer and trimer were also prepared. Looking specifically at the α-proton region (ca. 3.5 to 5.5ppm), we can see there is a strong similarity between the sample peaks and the peaks seen for the trimer standard (shown inset). Looking then at the amide proton region (ca. 7 to 11ppm) for the sample and all standards, no peak could be observed for the monomer (as expected), while peaks could be observed for both the dimer and trimer. The sample showed a number of peaks in this region (as would be expected for a complex mixture of oligomers and side products), but peaks in the right regions for the dimer and trimer could be seen. This shows that the proton NMR is not at all inconsistent with the presence of amide bonds.

5.2.2.9 Determination of soluble oligomer yields

The concentration of the smaller soluble oligomer products (1-mer to 6-mer; also DKP) was established by integration of absorbance values (195nm), and calibration with commercially-available standards. The calibration constants for larger oligomers were estimated based on the mean absorbance per glycine unit in the larger standards, which was observed to become approximately constant > 3-mer. (See Figure A4).

The yield was then calculated as a proportion of glycine (or glycine oligomer) starting material input. The data is averaged over three repetitions and error bars show the standard deviation.

5.2.2.10 RP-HPLC for amino acid analysis

Reversed-Phase HPLC analysis was performed using an Agilent 1260 HPLC system fitted with a reversed-phase C18 column (Poroshell HPH-C18, 300mm, 2.7µm), guard-column (UHPLC Guard 3PK, Poroshell HPH-C18, 15mm, 2.7µm). Samples were injected in 1µL aliquots and eluted at 0.620ml·min^{-1} constant flow rate with changing the gradient between aqueous mobile phase consisting of 10mM Na_2HPO_4, 10mM $Na_2B_4O_7$ and 5mM NaN_3 adjusted to pH=8.2 using HCl and organic mobile phase consisting of acetonitrile:methanol:water (45:45:10, v:v:v). Co-oligomerisation products were detected at 338 nm and 262 nm and assigned by their retention time previously determined when running individual amino acid monomer samples. The instrument was controlled and the resulting data analysed using Agilent Technologies OpenLAB Software.

Vial Number	**Label**
1	Injector diluent
2	Borate buffer
3	OPA
4	FMOC

Table 5: Vial number positions for the derivatisation reaction reagents.

Time (min)	Aqueous Mobile Phase (%)	Organic Mobile Phase (%)
0.35	98.0	2.0
13.40	43.0	57.0
13.50	0.0	100.0
20.00	98.0	2.0

Table 6: Elution gradient.

5.2.2.10.1 Auto-sampler sample preparation

1. Wash needle in flushport for 10 s
2. Draw 2.5 µL from location "Vial 2" with default speed using default offset
3. Draw 1 µL from sample with default speed using default offset
4. Mix 3.5 µL from seat with default speed for 5 times
5. Wait 0.2 min
6. Draw 0.5 µL from location "Vial 3" with default speed using default offset
7. Mix 4 µL from seat with default speed for 10 times
8. Draw 0.4 µL from location "Vial 4" with default speed using default offset
9. Mix 4.4 µL from seat with default speed for 10 times
10. Draw 32 µL from location "Vial 1" with default speed using default offset
11. Mix 20 µL from seat with maximum speed for 2 times
12. Inject
13. Wait 0.1 min
14. Switch valve to "Bypass"
15. Wash needle in flushport for 10 s

5.2.3 APS system description

5.2.3.1 Apparatus

The synthesis took place in bespoke automated apparatus comprising a set of programmable syringe pumps (C3000 Tricontinent), that were employed to flow the solutions to the heated reaction vessels. The pumps were controlled employing in-house developed LabView applications. Standard PEEK fittings were used to connect the tubing (1/16“OD, (0.3 mm ID)) and the reaction vessels. Custom 3D-printed caps / fittings were manufactured for the reaction vessels. Two different contact hotplates (RCT basic, IKA) fitted with DrySyn heat-transfer blocks were used in parallel to perform an array of experiments under different temperature conditions.

5.2.3.2 Operation and program

The unit operations of each cycle are controlled by LabView. We designed a program capable of controlling the flow rate of addition, volume and cycle time of each individual experiment within a continuous loop. Each one of these values can be easily modified between experiments, as well as the total number of cycles. Before starting the experiment, the desired values were entered in the program, and the experiments were left running autonomously over a specific period of time (depending on the number of cycles).

A typical peptide synthesis experiment involved: a) The preparation of a dilute solution of starting material was performed by taking an aliquot of 350µl from a 1 M solution of glycine, to which we subsequently added: 1ml of a 1M NaCl solution, 2.55ml of (HPLC) water, and finally the pH was adjusted to 9.8 by adding 100µl of 1M NaOH. b) Glass reaction vessels were placed in the corresponding Drysyn hotplate inserts. Custom-made 3D printed polypropylene lids with integrated holes were placed on each vial to connect them to the pumps, which facilitated the evaporation during the drying step. Water solutions were connected to each individual pump to deliver a given volume in each re-hydrating step. c) Then, the prepared solution was injected in a pre-heated vial (T=130 °C). d) The process inputs (volume, flow rate, dehydration time and number of cycles) were entered for each pump or set of pumps. The pumps were initialised and tested before starting the experiment to ensure their correct functioning. Finally, the array of experiments was started by pressing the START button.

e) After initialising the program, the vials were kept at 130 °C for a given time, in order to evaporate the solution to complete dryness. Once a cycle was finished, the process was re-started by re-hydrating the sample with 4ml of HPLC water. f) Once finished, products were collected for analysis by adding 8ml of a 0.1% TFA aqueous solution. Then, 500µl of the extracted sample were taken for HPLC analysis.

Although the reactions shown here were performed in glass vials, we note that the same series of oligomeric products were also observed if the reactions were run in Teflon® reactors (see Figure A7). While NaCl is present in these reactions; omission of NaCl in otherwise identical reactions resulted in no significant drop in yield. Furthermore, while NaOH was used to adjust pH, reactions using LiOH instead show that Na^+ is not crucial.

5.2.3.3 LabView Control Software

The control over the fluids described in the manuscript is implemented with Tricontinent syringe pumps operated through a USB to Serial connection to a laptop (Dell Latitude 3440). We developed a user interface for simple control of the pumps as well as easier implementation of the experimental design through National Instruments LabVIEW 2013 SP1 (32 bit). Figs. A1&A3 show screen shots of the main control software, "pump_control.vi," as well as an important Sub VI, "Pick up and Deliver (dist vale).vi". Commands are sent with the inbuilt VISA functions included with a LabVIEW installation. The Tricontinent C-Series Precision Pump (C3000/C24000) Software Manual and the C3000 Precision Pump Operator's Manual (available from the Tricontinent website, http://www.tricontinent.com) were used to inform the writing this control software.

5.2.4 APS reactions and sample preparation

5.2.4.1 Dehydration time dependence experiments

Starting from an aqueous solution of glycine monomer and keeping the reaction at pH=9.75, we show the conversion to oligomers depending on the dehydration time (after 1, 3, 5, 7, 15 and 24 h) at three different temperatures (90, 110 and 130ºC). The highest conversion takes places after 15h.

In all experiments, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was mixed with 1ml of a 1M solution of NaCl + 2.55ml HPLC water at pH=9.75 (adjusted by adding 100µl of NaOH 1M) .

Each sample was dehydrated in a single cycle for different periods of time: 1, 3, 5, 7, 15, 24, 48, 72 and 96h. Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution.

5.2.4.2 pH dependence experiments

Starting from an aqueous solution of glycine monomer and keeping the reaction at the same temperature (130 °C) for a dehydrating time of 24h, we show the conversion to oligomers depending on the pH (between pH=10.0 and 2.15).

In all experiments, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was mixed with 1ml of a 1M solution of NaCl + 2.55ml HPLC water at various pH values (adjusted by adding a 1M NaOH or 1M HCl solution) . Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution.

5.2.4.3 Cycle dependency experiments

Starting from an aqueous solution of glycine monomer and keeping the reaction at the same temperature (130 °C) and pH= 2.61 for a dehydrating time of 24 hours, we show the conversion to oligomers depending on number of hydration-dehydration cycles (*n*=1-3).

In all experiments, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was mixed with 1ml of a 1M solution of NaCl + 2.60ml HPLC water, and adjusted to pH=2.61 by adding 50µl of a 1M HCl solution.
Each sample was dehydrated and subsequently rehydrated for a given number of cycles. Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4 ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution.

5.2.4.4 Glycinamide and DKP Oligomerisation Reactions

In this experiment, DKP (0.0014g) was dissolved in 1 ml of a 1M solution of NaCl + 2.55ml HPLC water, and adjusted to pH=9.75 using 1M NaOH. The sample was left to dehydrate for 24h in a single cycle dehydration experiment. Once finished, products were prepared for analysis by dissolution in 8ml of a 0.1% TFA aqueous solution.
In this experiment, Glycinamide (0.0014g) was dissolved in 1 ml of a 1M solution of NaCl + 2.5ml HPLC water, and adjusted to pH=9.75 using 1M NaOH. The sample was left to dehydrate for 24h in a single cycle dehydration experiment. Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution.

5.2.4.5 Glycine oligomerisation in a Teflon® vessel

In this experiment, glycine (350µl from a 1M Gly solution) was mixed with 1ml of a 1M solution of NaCl + 2.55ml HPLC water, and adjusted to pH=9.75 using 1M NaOH. The sample was left to dehydrate for 24h in a single cycle dehydration experiment in a Teflon® vessel. Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution. (See Figure A5).

5.2.4.6 Evidence for concurrent bond breakage and formation

In this experiment, Glycylglycine (175µl from a 1M GlyGly solution, so $1.75 \cdot 10^{-4}$ moles) was mixed with 1ml of a 1M solution of NaCl + 2.725ml HPLC water, and adjusted to pH=9.8 by adding 100 µl of a 1M NaOH solution. The reaction was left to dehydrate at a) 110 ºC and b) 130ºC for a single dehydration cycle. Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution.
Since $(Gly)_n$ series are observed with odd-numbered n (i.e. n = 3,5, 7, etc), we can conclude that peptide bonds must have been both broken and formed in the same experimental cycle. We observe a greater occurrence of this at higher temperatures.

5.2.4.7 Experiment using natrolite mineral and glycine monomer

Natrolite is a tectosilicate mineral species belonging to the zeolite group. It is a basic hydrated sodium and aluminium silicate with the formula $Na_2Al_2Si_3O_{10} \cdot 2H_2O$. Natrolite was obtained from Richard Tayler Minerals, Cobham, Surrey, England, and was used (crushed) without further purification.
In this experiment, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was dissolved in 4ml water and added to ca. 500mg crushed natrolite. The sample was dehydrated in a single cycle for 24h at 130ºC. All the dried products were extracted by adding 4ml of HPLC water.

5.2.4.8 Hetero-oligomer synthesis

In all experiments, glycine (3.5ml from a 0.1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was mixed with 3.5ml of a 0.1M solution of a second amino acid (A, D, E, K, H, P, V and T). Then, 1ml of a 1M solution of NaCl was added, and the pH was adjusted to pH = 3 using a small amount of 1M HCl solution. All samples were dehydrated at 130 °C for 15h; they were rehydrated (4ml of HPLC water) and dehydrated (again, 130 °C for 15h) to achieve a total number of 5 dehydration-hydration cycles. Once finished, products were prepared for analysis by dissolution in 8ml of water.

5.2.4.9 Confirmation of the production of hetero-oligomers incorporating different amino acids

Since increased complexity results from the co-condensation of multiple amino acid monomers and makes thorough characterization of all the resulting products (presuming a relatively uncontrolled/non-selective reaction) more challenging, a "virtual" BPC, constructed. This is derived from EICs resulting from a combinatorial list of the all possible oligomer-forming combinations of both/all the amino acids (i.e. only species incorporating all starting amino acids are represented, not homo-oligomers. In all cases many peaks are observed in these traces, suggesting the formation of many co-oligomerisation products.

Secondly, some representative components of these complex mixtures were chosen for MS/MS analysis to corroborate the peptide nature of the products observed.

5.2.4.10 NaCl concentration dependence (I)

We show that conversion to oligomers does not depend strongly on the concentration of NaCl added to the starting solution, with a series of experiments. Each experiment started from an aqueous solution of glycine monomer, and was maintained at 130 °C and pH=9.75 for a dehydrating time of 15h.

In all experiments, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was diluted in a total volume of 4ml. Different volumes of a 5M NaCl solution were added to adjust the NaCl concentration to 0, 0.025, 0.25 and 1M. The pH was adjusted to 9.75 by adding 100µl of a 1M NaOH solution (except in one case, where LiOH was used, to demonstrate

that Na^+ was not vital). Each sample was dehydrated for one cycle. Once finished, products were prepared for analysis by dissolution of the dehydrated samples in 4ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution.

5.2.4.11 NaCl concentration dependence (II)

In all experiments, glycine (350µl from a 1M Gly solution, so $3.5 \cdot 10^{-4}$ moles) was diluted in a total volume of 4ml, and prepared at various pH values (adjusted by adding a 1M NaOH or 1M HCl solution). The samples were prepared in duplicates to compare the effect of NaCl in our reaction. Therefore, 1ml of a 1M NaCl solution was added to one of the duplicate samples. All reactions were left to dehydrate on a hotplate at 130 °C for 15h. The samples adjusted to basic pH were prepared for analysis by dissolution of the dehydrated samples in 4 ml of a 0.1% TFA aqueous solution. Then, 100µl of this solution were added to 100µl of a 0.05M H_3PO_4 solution. The samples adjusted to acidic pH were prepared for analysis by dissolution in 8ml of H_2O.

5.2.4.12 Controlling the rate of evaporation experiments.

In this experiment, we wanted to study the oligomerisation of glycine monomer as a function of the evaporation rate. This was accomplished by performing two reactions in parallel that contained the exact same starting solution: glycine (3500µl from a 0.1M Gly solution, so $3.5 \cdot 10^{-4}$ moles was diluted in a total volume of 4ml, and prepared at various pH values, adjusted by adding a 1M NaOH or 1M HCl solution), but within different vials: a 3D printed plastic lid with a 2.5mm hole was placed on one of the vials (slow evaporation rate), whilst the other vial was left open (fast evaporation rate). Both reactions were left to dehydrate on a hotplate at 130 °C for 15h. Once finished, the samples adjusted to basic pH were prepared for analysis by dissolution of the dehydrated samples in 8 ml of a 0.1M H_3PO_4 aqueous solution. The samples adjusted to neutral or acidic pH were prepared for analysis by dissolution in 8ml of H_2O.

5.2.4.13 Glycine oligomerisation with different acids and bases

In this set of experiments, we wanted to the study the effect of using different acids or bases in a standard glycine oligomerisation reaction. In this case, the rate of evaporation was also studied by performing two reactions in parallel that contained the exact same starting solution: glycine (3500µl from a 0.1M Gly solution, so $3.5 \cdot 10^{-4}$ moles was diluted in a total volume of 4ml, and prepared at various pH values, adjusted by adding the corresponding amount of: NaOH, KOH, LiOH, HCl, H_3PO_4 or H_2SO_4 1M solution). Both reactions were left to dehydrate on a hotplate at 130 °C for 15h. Once finished, the samples adjusted to basic pH were prepared for analysis by dissolution of the dehydrated samples in 8 ml of a 0.1M H_3PO_4 aqueous solution. The samples adjusted to neutral or acidic pH were prepared for analysis by dissolution in 8ml of H_2O.

5.2.4.14 Glycine oligomerisation upon the addition of different amounts of H_3PO_4 and H_2SO_4

In this set of experiments, we wanted to the study whether adding higher amounts of acid (H_3PO_4 or H_2SO_4) in a standard glycine oligomerisation reaction contributed to an increase in yield. Glycine (3500µl from a 0.1M Gly solution, so $3.5 \cdot 10^{-4}$ moles was diluted in a total volume of 4ml, and prepared at various pH values, adjusted by adding increasing amounts of a H_3PO_4 or H_2SO_4 2M solution: 25, 50, 100 and 250 µl ($5 \cdot 10^{-5}$, $1 \cdot 10^{-4}$, $2 \cdot 10^{-4}$ and $5 \cdot 10^{-4}$ moles, respectively).

All reactions were left to dehydrate on a hotplate at 130 °C for 15h. Once finished, the samples were prepared for analysis by dissolution in 8ml of H_2O.

5.2.4.15 Glycine-Valine co-oligomerisation upon the addition of different amounts of H_3PO_4 and H_2SO_4

In this set of experiments, we wanted to the study whether adding higher amounts of acid (H_3PO_4 or H_2SO_4) in a glycine-valine co-oligomerisation reaction contributed to an increase in yield. Glycine (1750µl from a 0.1M Gly solution, so $1.75 \cdot 10^{-4}$ moles) and Valine (1750µl from a 0.1M Val solution, so $1.75 \cdot 10^{-4}$ moles) were diluted in a total volume of 4ml, and prepared at various pH values, adjusted by adding increasing amounts

of a H_3PO_4 or H_2SO_4 2M solution: 25, 50, 100 and 250 µl ($5 \cdot 10^{-5}$, $1 \cdot 10^{-4}$, $2 \cdot 10^{-4}$ and $5 \cdot 10^{-4}$ moles, respectively).

All reactions were left to dehydrate on a hotplate at 130 °C for 15h. Once finished, the samples were prepared for analysis by dissolution in 8ml of H_2O.

5.2.4.16 Co-oligomerisation matrix experiments

In this set of experiments, we studied the homo- and hetero-oligomerisation between 10 different amino acids. The co-oligomerisation reactions were also studied in the presence of 3 different acids (HCl, H_3PO_4 and H_2SO_4) and one base (NaOH). For each set of reactions, a standard 0.1M solution was prepared for each individual amino acids, and adjusted to pH=2.2 (using the acids mentioned above) or pH=9.8 (for NaOH).

The amino acids used were: Glycine, L-Alanine, L-Leucine, L-Aspartic Acid, L-Glutamic Acid, L-Lysine, L-Histidine, L-Valine, L-Proline and L-Threonine.

In the case of L-Aspartic acid and L-Glutamic acid, the standard 0.1M solutions were placed in the ultrasonicator at 65 °C for at least 30 minutes to ensure their complete dissolution.

In the homo-oligomerisation reactions, 3.5ml of a 0.1M solution ($3.5 \bullet 10^{-4}$ moles) of a given amino acid were dehydrated in a hotplate at 130 °C for 15h.

In the hetero-oligomerisation reactions, 1.75ml of a 0.1M solution from amino acid 1 ($1.75 \bullet 10^{-4}$ moles) were mixed with 1.75ml of a 0.1M solution from amino acid 2 ($1.75 \bullet 10^{-4}$ moles) and dehydrated in a hotplate at 130 °C for 15h.

Once finished, the samples were prepared for analysis by dissolution in 8ml of H_2O.

5.2.4.17 Hotplate calibration

Co-oligomerisation experiments were performed in an in-built hotplate. The hotplate was comprised of an aluminium heating block with 75 holes to fit 7 ml vials (23 mm in diameter) and a 3 mm thick aluminium plate with a bespoke wire-wound silicone heater. Temperature was controlled via a PID controller and measured with a thermocouple.

The temperature was measured at different positions across the plate with a thermocouple attached to a *Hannah Instrument*© thermocouple thermometer submerged in a vial filled with mineral oil when the hotplate temperature was set to 130 °C.

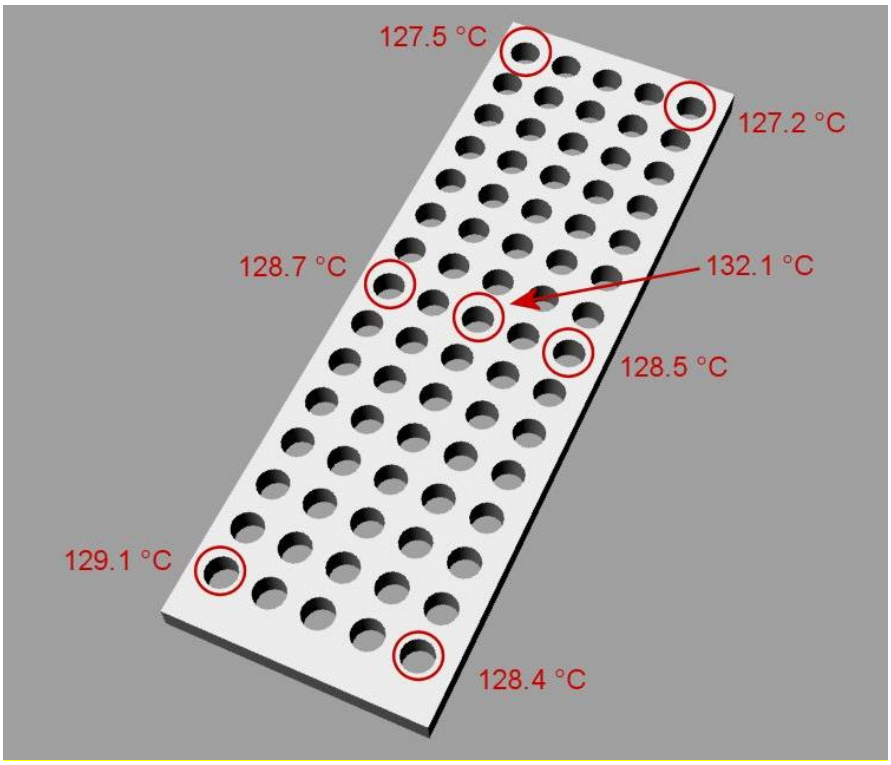


Figure 113: Different measured temperatures obtained when a thermocouple was submerged in a vial filled with mineral oil, and the hotplate temperature was set to 130 °C.

Different temperature recordings showed that the temperature slightly decreased on the edges of the hotplate. The average temperature was 128.78 ± 1.48 °C.

5.2.5 Environmental programming of peptide synthesis

5.2.5.1 Order of addition experiments

In this set of experiments, three different amino acid solutions were added in a different order of addition under successive dehydration-hydration cycles.

1. Three individual solutions of amino acids were prepared to a final concentration of 0.1 M and adjusted to pH=2.5 by adding HCl.

   Amino acid 1: glycine

   Amino acid 2: L-alanine

   Amino acid 3: L-histidine.

2. The order in which the different amino acid solutions were added was as the following:

	Experiment						
Cycle	1	2	3	4	5	6	7
1	**Gly**	**Gly**	**Ala**	**Ala**	**His**	**His**	**Gly+Ala+His**
2	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
3	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
4	**Ala**	**His**	**Gly**	**His**	**Gly**	**Ala**	**Gly+Ala+His**
5	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
6	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
7	**His**	**Ala**	**His**	**Gly**	**Ala**	**Gly**	**Gly+Ala+His**
8	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
9	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O

3. 3.5ml of a 0.1M solution of each amino acid ($3.5 \cdot 10^{-4}$ moles) were added in cycles 1, 4 and 7.
4. In the experiment where a mixture of the three amino acid solutions was added together, 1.16ml of each amino acid solution was added.
5. 3.5 ml of HPLC water were added in cycles 2,3,5,6,8 and 9.
6. Each single dehydration-hydration cycle was performed on a hotplate at 130C for 12h.
7. Once finished, all the samples were diluted by adding 6 ml of HPLC water.

8. 500 µl were taken for LC-MS analysis. The remaining sample was dialysed with a G2 Float-a-lyser (500-1000Da) cut-off (5ml) for 20h.
9. Once the dialysis was completed, the samples were left to freeze-dry for 48h.
10. An exact amount of solid peptide material was weighed in a micro-balance, and diluted with the right amount of water to achieve a final concentration of 2mg/ml.
11. Once the peptide stock solutions were prepared, an aliquot was taken for esterase activity assay.

5.2.5.2 Soluble salts experiments

In this set of experiments, one solution containing an equimolar amount of three different amino acids was dehydrated in the presence of different soluble salts under successive dehydration-hydration cycles.

1. A solution containing three different amino acids (Gly, Ala, His) was prepared to a final concentration of 0.033 M (each) and adjusted to pH=2.5 by adding HCl.
2. 7 different soluble salt solutions were prepared at a final concentration of 1 M.

	Experiment						
	1	2	3	4	5	6	7
Cycle	NaCl	KCl	LiCl	NH_4Cl	$MgCl_2$	$CuCl_2$	$EuCl_3$
1	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**
2	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
3	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
4	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**
5	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
6	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
7	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**
8	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
9	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O

3. 1ml of a 1M soluble salt solution was added in cycle 1 to each individual experiment.
4. 3.5ml of the amino acids solutions were added in cycles 1, 4 and 7.
5. 3.5 ml of HPLC water were added in cycles 2,3,5,6,8 and 9.
6. Each single dehydration-hydration cycle was performed on a hotplate at 130C for 12h.
7. Once finished, all the samples were diluted by adding 6 ml of HPLC water.

8. 500µl were taken for LC-MS analysis. The remaining sample was dialysed with a G2 Float-a-lyser (500-1000Da) cut-off (5ml) for 20h.
9. Once the dialysis was completed, the samples were left to freeze-dry for 48h.
10. An exact amount of solid peptide material was weighed in a micro-balance, and diluted with the right amount of water to achieve a final concentration of 2mg/ml.
11. Once the peptide stock solutions were prepared, an aliquot was taken for esterase activity assay.

5.2.5.3 Insoluble salts experiments

In this set of experiments, one solution containing an equimolar amount of three different amino acids was dehydrated in the presence of different insoluble salts under successive dehydration-hydration cycles.

1. A solution containing three different amino acids (Gly, Ala, His) was prepared to a final concentration of 0.033 M (each) and adjusted to pH=2.5 by adding HCl.
2. 7 different insoluble salt solutions were prepared at a final concentration of 1 M.

	Experiment						
	1	2	3	4	5	6	7
Cycle	Alumina	Montmorillonite	Mica	Goethite	Quartz	Natrolite	Silica
1	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**
2	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
3	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
4	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**
5	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
6	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
7	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**	**Gly+Ala+His**
8	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
9	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O

3. 0.2g of an insoluble salt were added in cycle 1 to each individual experiment.
4. 3.5ml of the amino acids solutions were added in cycles 1, 4 and 7.
5. 3.5 ml of HPLC water were added in cycles 2,3,5,6,8 and 9.
6. Each single dehydration-hydration cycle was performed on a hotplate at 130C for 12h.

7. Once finished, all the samples were diluted by adding 6 ml of HPLC water.
8. 500µl were taken for LC-MS analysis. The remaining sample was dialysed with a G2 Float-a-lyser (500-1000Da) cut-off (5ml) for 20h.
9. Once the dialysis was completed, the samples were left to freeze-dry for 48h.
10. An exact amount of solid peptide material was weighed in a micro-balance, and diluted with the right amount of water to achieve a final concentration of 2mg/ml.
12. Once the peptide stock solutions were prepared, an aliquot was taken for esterase activity assay.

5.2.5.4 Peptide stock solution preparation

0.5mg/ml stock solutions were prepared from previously lyophilised peptides. Once the stock solutions were prepared, they were used immediately to perform the assay.

5.2.5.5 p-Nitrophenyl acetate assay

A buffered substrate solution was prepared by adding 300µl of p-Nitrophenyl acetate 0.1M (in acetonitrile) and 300µl of HEPES buffer 1M to 11400µl of water. The final amount of acetonitrile present was 1.875%.

150µl of buffered substrate solution were then added to 50µl of peptide stock solution (being the final substrate concentration 1.875mM). The approximate final concentration of the peptide stock solution was 0.21875mM.

Kinetic measurements were performed in an Infinite M200Pro Tecan plate reader monitoring the absorbance of the p-nitrophenyl acetate hydrolysis product (p-nitrophenol) at 405 nm and at 25 °C, in 96-well plate every 5 min for 2h.

5.2.5.5.1 Amino acid monomer control experiment

As a control experiment, the p-NPA assay was run for a solution containing an equimolar mixture of Gly, Ala and His monomers. First, a 0.5 mg/ml solution (1 ml) was prepared for each individual amino acid. The three amino acid solutions were then mixed, giving a final concentration of 0.1667 mg/ml for each amino acid. The p-NPA assay was performed as described in section 5.2.5.5.

6 References

(1) Bell, E. A.; Boehnke, P.; Harrison, T. M.; Mao, W. L. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *Early Edit* (20), 1–4.

(2) Mojzsis, S. J.; Arrhenius, G.; McKeegan, K. D.; Harrison, T. M.; Nutman, a P.; Friend, C. R. *Nature* **1996**, *384* (6604), 55–59.

(3) Joyce, G. F. *PLoS Biol.* **2012**, *10* (5).

(4) Chyba, C.; Sagan, C. *Nature* **1992**, *355* (6356), 125–132.

(5) Woese, C. R. *Science (80-.).* **1964**, *144* (2), 1030–1031.

(6) Crick, F. H. C. *J. Mol. Biol.* **1968**, *38* (3), 367–379.

(7) Barrell, B. G.; Bankier, a T.; Drouin, J. *Nature* **1979**, *282* (5735), 189–194.

(8) Knight, R. D.; Freeland, S. J.; Landweber, L. F. *Nat. Rev. Genet.* **2001**, *2* (1), 49–58.

(9) Hao, B. *Science (80-.).* **2002**, *296* (5572), 1462–1466.

(10) Woese, C. R.; Kandler, O.; Wheelis, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (12), 4576–4579.

(11) Hug, L. A.; Baker, B. J.; Anantharaman, K.; Brown, C. T.; Probst, A. J.; Castelle, C. J.; Butterfield, C. N.; Hernsdorf, A. W.; Amano, Y.; Kotaro, I.; Suzuki, Y.; Dudek, N.; Relman, D. A.; Finstad, K. M.; Amundson, R.; Thomas, B. C.; Banfield, J. F. *Nat. Microbiol.* **2016**, *1* (May), 16048.

(12) Woese, C. R. *Proc. Natl. Acad. Sci.* **2000**, *97* (15), 8392–8396.

(13) Podolsky, S. H.; Tauber, A. I. *Q. Rev. Biol.* **1994**, *69* (2), 253–254.

(14) Garrison, W. M.; Morrison, D. C.; Hamilton, J. G.; Benson, A. A.; Calvin, M. *Science (80-.).* **1951**, *114* (2964), 416–418.

(15) Miller, S. L. *Science* **1953**, *117* (3046), 528–529.

(16) Miller, S. L. *Biochim. Biophys. Acta* **1957**, *23* (3), 480–489.

(17) Oró, J. *Nature* **1961**, *191*, 1193.

(18) ORÓ, J.; KAMAT, S. S. *Nature* **1961**, *190* (4774), 442–443.

(19) ORÓ, J. *Nature* **1961**, *190* (4774), 389–390.

(20) Cech, T. R. *Biosci. Rep.* **2004**, *24* (4-5), 362–385.

(21) Altman, S. *Biosci Rep* **1990**, *10* (4), 317–337.

(22) Yarus, M. *Cold Spring Harb. Perspect. Biol.* **2011**, *3* (4), 1–8.

(23) Higgs, P. G.; Lehman, N. *Nat. Rev. Genet.* **2014**, *16* (1), 7–17.

(24) Cech, T. R. *RNA Worlds* **2010**, 9–13.

(25) Wächtershäuser, G. *Microbiol. Rev.* **1988**, *52* (4), 452–484.

(26) Cody, G. D. *Science (80-.).* **2000**, *289* (5483), 1337–1340.

(27) Huber, C.; Wachtershauser, G. *Science (80-.).* **2006**, *314* (5799), 630–632.

(28) Huber, C.; Wächtershäuser, G. *Science* **1998**, *281* (5377), 670–672.

(29) Martin, W.; Russell, M. J. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2003**, *358* (1429), 59–83; discussion 83–85.

(30) Martin, W.; Baross, J.; Kelley, D.; Russell, M. J. *Nat. Rev.* **2008**, *6* (11), 805–814.

(31) Martin, W.; Russell, M. J. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2007**, *362* (1486), 1887–1925.

(32) Ikehara, K.; Omori, Y.; Arai, R.; Hirose, A. *J. Mol. Evol.* **2002**, *54* (4), 530–538.

(33) Ikehara, K. *Orig. Life Evol. Biosph.* **2015**.

(34) Oba, T.; Fukushima, J.; Maruyama, M.; Iwamoto, R.; Ikehara, K. *Orig. Life Evol. Biosph.* **2005**, *35*, 447–460.

(35) Fox, S. W.; HARADA, K. *Science (80-.).* **1958**, *128* (3333), 1214–1214.

(36) Segré, D.; Ben-Eli, D.; Deamer, D. W.; Lancet, D. *Orig. Life Evol. Biosph.* **2001**, *31* (1-2), 119–145.

(37) Borsenberger, V.; Crowe, M. A.; Lehbauer, J.; Raftery, J.; Helliwell, M.; Bhutia, K.; Cox, T.; Sutherland, J. *Chem. Biodivers.* **2004**, *1* (2), 203–246.

(38) Pascal, R.; Boiteau, L.; Commeyras, a. *Prebiotic Chem.* **2005**, No. August, 69–122.

(39) Strecker, A. *Ann. der Chemie und Pharm.* **1850**, *75* (1), 27–45.

(40) Jakschitz, T. a E.; Rode, B. M. *Chem. Soc. Rev.* **2012**, *41* (16), 5484–5489.

(41) Rode, B. M.; Fitz, D.; Jakschitz, T. *Chem. Biodivers.* **2007**, *4* (12), 2674–2702.

(42) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Science* **2008**, *322* (5900), 404.

(43) Parker, E. T.; Zhou, M.; Burton, A. S.; Glavin, D. P.; Dworkin, J. P.; Krishnamurthy, R.; Fern??ndez, F. M.; Bada, J. L.; Fernández, F. M.; Bada, J. L. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (31), 8132–8136.

(44) Taillades, J.; Beuzelin, I.; Garrel, L.; Tabacik, V.; Bied, C.; Commeyras, a. *Orig. Life Evol. Biosph.* **1998**, *28*, 61–77.

(45) Sephton, M. A. *Nat. Prod. Rep.* **2002**, *19* (3), 292–311.

(46) Schmitt-Kopplin, P.; Gabelica, Z.; Gougeon, R. D.; Fekete, A.; Kanawati, B.; Harir, M.; Gebefuegi, I.; Eckel, G.; Hertkorn, N. *Proc Natl Acad Sci U S A* **2010**, *107* (7), 2763–2768.

(47) Kuan, Y.; Charnley, S. B.; Huang, H.; Tseng, W.; Kisiel, Z. *Astrophys. J.* **2003**, *593* (2), 848–867.

(48) Muñoz Caro, G. M.; Meierhenrich, U. J.; Schutte, W. a; Barbier, B.; Arcones Segovia, a; Rosenbauer, H.; Thiemann, W. H.-P.; Brack, a; Greenberg, J. M. *Nature* **2002**, *416* (6879), 403–406.

(49) Bernstein, M. P.; Dworkin, J. P.; Sandford, S. A.; Cooper, G. W.; Allamandola, L. J. *Nature* **2002**, *416*, 401–403.

(50) Meinert, C.; Filippi, J. J.; De Marcellus, P.; Le Sergeant D'Hendecourt, L.; Meierhenrich, U. J. *Chempluschem* **2012**, *77* (3), 186–191.

(51) Wakamiya, M.-. **2001**, *27* (2), 207–215.

(52) Martins, Z.; Price, M. C.; Goldman, N.; Sephton, M. a; Burchell, M. J. *Nat. Geosci* **2013**, *6* (12), 1045–1049.

(53) Sanchez, R. A. A.; Ferris, J. P. P.; Orgel, L. E. E. *Science (80-.).* **1966**, *154* (3750), 784–785.

(54) Saladino, R.; Botta, G.; Pino, S.; Costanzo, G.; Di Mauro, E. *Chem. Soc. Rev.* **2012**, *41* (16), 5526–5565.

(55) Breslow, R. *Tetrahedron Lett.* **1959**, *1* (21), 22–26.

(56) Gabel, N. W.; Ponnamperuna, C. *Nature* **1967**, *216* (5114), 453–455.

(57) Reid, C.; Orgel, L. E. *Nature* **1967**, *216* (5114), 455–455.

(58) Schwartz, A. W.; de Graaf, R. M. *J. Mol. Evol.* **1993**, *36* (2), 101–106.

(59) Benner, S. A.; Kim, H. J.; Carrigan, M. A. *Acc. Chem. Res.* **2012**, *45* (12), 2025–2034.

(60) Ricardo, A. *Science (80-.).* **2004**, *303* (5655), 196–196.

(61) Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Biol.* **1972**, *67* (1), 25–33.

(62) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *459* (7244), 239–242.

(63) Becker, S.; Thoma, I.; Deutsch, A.; Gehrke, T.; Mayer, P.; Zipse, H.; Carell, T. *Science (80-.).* **2016**, *352* (6287), 833–836.

(64) Joyce, G. F.; Schwartz, a W.; Miller, S. L.; Orgel, L. E. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4398–4402.

(65) Kolb, V. M.; Dworkin, J. P.; Miller, S. L. *J. Mol. Evol.* **1994**, *38* (6), 549–557.

(66) Hud, N. V. N.; Cafferty, B. J. B.; Krishnamurthy, R.; Williams, L. L. D. *Chem. Biol.* **2013**, *20* (4), 466–474.

(67) Cafferty, B. J.; Hud, N. V. *Curr. Opin. Chem. Biol.* **2014**, *22C*, 146–157.

(68) Cafferty, B. J.; Fialho, D. M.; Khanam, J.; Krishnamurthy, R.; Hud, N. V. *Nat. Commun.* **2016**, *7*, 11328.

(69) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. *Nat. Chem.* **2015**, *7* (4), 301–307.

(70) Costanzo, G.; Saladino, R.; Crestini, C.; Ciciriello, F.; Di Mauro, E. *BMC Evol. Biol.* **2007**, *7 Suppl 2*, S1.

(71) Saladino, R.; Crestini, C.; Pino, S.; Costanzo, G.; Di Mauro, E. *Phys. Life Rev.* **2012**, *9* (1), 84–104.

(72) Barks, H. L.; Buckley, R.; Grieves, G. A.; Di Mauro, E.; Hud, N. V.; Orlando, T. M. *ChemBioChem* **2010**, *11* (9), 1240–1243.

(73) Šponer, J. E.; Šponer, J.; Nováková, O.; Brabec, V.; Šedo, O.; Zdráhal, Z.; Costanzo, G.; Pino, S.; Saladino, R.; Di Mauro, E. *Chem. - A Eur. J.* **2016**, *22* (11), 3572–3586.

(74) Saladino, R.; Ciambecchini, U.; Crestini, C.; Costanzo, G.; Negri, R.; Di Mauro, E. *ChemBioChem* **2003**, *4* (6), 514–521.

(75) Saladino, R.; Carota, E.; Botta, G.; Kapralov, M.; Timoshenko, G. N.; Rozanov, A. Y.; Krasavin, E.; Di Mauro, E. *Proc. Natl. Acad. Sci.* **2015**, No. 48, 201422225.

(76) Bruce Martin, R. *Biopolymers* **1998**, *45* (5), 351–353.

(77) Shock, E. L. *Geochim. Cosmochim. Acta* **1992**, *56* (9), 3481–3491.

(78) Meggy, A. B. *J. Chem. Soc.* **1953**, *1* (1), 851–855.

(79) Meggy, A. B. *J. Chem. Soc.* **1956**, No. 1444, 1444.

(80) Harada, K.; Fox, S. W. *J. Am. Chem. Soc.* **1958**, *80* (11), 2694–2697.

(81) FOX, S. W. *Nature* **1964**, *201* (4917), 336–337.

(82) Fox, S. W.; Bahn, P. R.; Dose, K.; Harada, K.; Hsu, L.; Ishima, Y.; Jungck, J.; Kendrick, J.; Krampitz, G.; Lacey, J. C.; Matsuno, K.; Melius, P.; Middlebrook, M.; Nakashima, T.; Pappelis, A.; Pol, A.; Rohlfing, D. L.; Vegotsky, A.; Waehneldt, T. V.; Wax, H.; Yu, B. *J. Biol. Phys.* **1995**, *20* (1-4), 17–36.

(83) Nakashima, T.; Fox, S. W. *Biosystems* **1981**, *14* (2), 151–161.

(84) Rohlfing, D. L. *Nature* **1967**, *216* (5116), 657–659.

(85) Rohlfing, D. L. *Science* **1970**, *169* (949), 998–1000.

(86) Rohlfing, D. L. *Science* **1974**, *6406* (4247), 5–7.

(87) Dose, K. *Orig. Life* **1974**, *5* (1-2), 239–252.

(88) Andini, S.; Benedetti, E.; Ferrara, L.; Paolillo, L.; Temussi, P. A. In *Cosmochemical Evolution and the Origins of Life*; Oró, J., Miller, S. L., Ponnamperuma, C., Young, R. S., Eds.; Springer Netherlands: Dordrecht, 1974; pp 147–153.

(89) Fox, S. W.; Harada, K. *Arch. Biochem. Biophys.* **1960**, *86*, 281–285.

(90) Varfolomeev, S. D.; Lushchekina, S. V. *Geochemistry Int.* **2014**, *52* (13), 1197–1206.

(91) Roweton, S.; Huang, S.; Swift, G. *J. Polym. Environ.* **1997**, *5* (3), 175–181.

(92) Liu, R.; Orgel, L. E. *Orig. Life Evol. Biosph.* **1998**, *28* (3), 245–257.

(93) Lambert, J.-F. *Orig. Life Evol. Biosph.* **2008**, *38* (3), 211–242.

(94) Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. *Chem. Rev.* **2014**, *114* (1), 285–366.

(95) Rode, B. M.; Schwendinger, M. G. *Orig. Life Evol. Biosph.* **1990**, *20* (5), 401–410.

(96) Griffith, E. C.; Vaida, V. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (39), 15697–15701.

(97) Dobson, C. M.; Ellison, G. B.; Tuck, A. F.; Vaida, V. *Proc. Natl. Acad. Sci.* **2000**, *97* (22), 11864–11868.

(98) Islam, M. N.; Kaneko, T.; Kobayashi, K. *Bull. Chem. Soc. Jpn.* **2003**, *76* (6), 1171–1178.

(99) Otake, T.; Taniguchi, T.; Furukawa, Y.; Kawamura, F.; Nakazawa, H.; Kakegawa, T. *Astrobiology* **2011**, *11* (8), 799–813.

(100) Ohara, S.; Kakegawa, T.; Nakazawa, H. *Orig. Life Evol. Biosph.* **2007**, *37* (3), 215–223.

(101) Imai, E.; Honda, H.; Hatori, K.; Brack, A.; Matsuno, K. *Science* **1999**, *283* (1984), 831–833.

(102) Tsukahara, H.; Imai, E. I.; Honda, H.; Hatori, K.; Matsuno, K. *Orig. Life Evol. Biosph.* **2002**, *32*, 13–21.

(103) Kawamura, K.; Takeya, H.; Kushibe, T.; Koizumi, Y. *Astrobiology* **2011**, *11* (5), 461–469.

(104) Cleaves, H. J.; Aubrey, a D.; Bada, J. L. *Orig. Life Evol. Biosph.* **2009**, *39* (2), 109–126.

(105) Sakata, K.; Kitadai, N.; Yokoyama, T. *Geochim. Cosmochim. Acta* **2010**, *74* (23), 6841–6851.

(106) Ponnamperuma, C.; Peterson, E. *Science (80-.).* **1965**, *147* (3665), 1572–1574.

(107) Eichberg, J.; Sherwood, E.; Epps, D. E.; Oró, J. *J. Mol. Evol.* **1977**, *10* (3), 221–230.

(108) Hawker, J. R.; Oró, J. *J. Mol. Evol.* **1981**, *17* (5), 285–294.

(109) Steinman, G.; Lemmon, R. M.; Calvin, M. *Science (80-.).* **1965**, *147* (3665), 1574–1575.

(110) Danger, G.; Michaut, A.; Bucchi, M.; Boiteau, L.; Canal, J.; Plasson, R.; Pascal, R. *Angew. Chemie - Int. Ed.* **2013**, *52* (2), 611–614.

(111) Ferris, J. P.; Hill, a R.; Liu, R.; Orgel, L. E. *Nature*. 1996, pp 59–61.

(112) Leuchs, H. *Berichte der Dtsch. Chem. Gesellschaft* **1906**, *39* (1), 857–861.

(113) Kricheldorf, H. R. *Angew. Chemie - Int. Ed.* **2006**, *45* (35), 5752–5784.

(114) Brack, A. *Orig. Life Evol. Biosph.* **1987**, *17* (3-4), 367–379.

(115) Brack, A. *Biosystems* **1982**, *15* (3), 201–207.

(116) Paecht-Horowitz, M.; Eirich, F. R. *Orig. Life Evol. Biosph.* **1988**, *18* (4), 359–387.

(117) Armstrong, D. W.; Seguin, R.; McNeal, C. J.; Macfarlane, R. D.; Fendler, J. H. *J. Am. Chem. Soc.* **1978**, *100* (14), 4605–4606.

(118) Commeyras, A.; Collet, H.; Boiteau, L.; Taillades, J.; Vandenabeele-Trambouze, O.; Cottet, H.; Biron, J. P.; Plasson, R.; Mion, L.; Lagrille, O.; Martin, H.; Selsis, F.; Dobrijevic, M. *Polym. Int.* **2002**, *51* (7), 661–665.

(119) Taillades, J.; Collet, H.; Garrel, L.; Beuzelin, I.; Boiteau, L.; Choukroun, H.; Commeyras, A. *J. Mol. Evol.* **1999**, *48* (6), 638–645.

(120) Collet, H.; Bied, C.; Mion, L.; Taillades, J.; Commeyras, A. *Tetrahedron Lett.* **1996**, *37* (50), 9043–9046.

(121) Danger, G.; Boiteau, L.; Cottet, H.; Pascal, R. *J. Am. Chem. Soc.* **2006**, *128* (23), 7412–7413.

(122) Danger, G.; Plasson, R.; Pascal, R. *Chem. Soc. Rev.* **2012**, *41* (16), 5416–5429.

(123) Leman, L.; Orgel, L.; Ghadiri, M. R. *Science* **2004**, *306* (5694), 283–286.

(124) Georgelin, T.; Jaber, M.; Bazzi, H.; Lambert, J.-F. *Orig. Life Evol. Biosph.* **2013**, *43* (4-5), 429–443.

(125) Martra, G.; Deiana, C.; Sakhno, Y.; Barberis, I.; Fabbiani, M.; Pazzi, M.; Vincenti, M. *Angew. Chemie - Int. Ed.* **2014**, *53* (18), 4671–4674.

(126) Mamajanov, I.; Macdonald, P. J.; Ying, J.; Duncanson, D. M.; Dowdy, G. R.; Walker, C. a.; Engelhart, A. E.; Fern??ndez, F. M.; Grover, M. A.; Hud, N. V.; Schork, F. J.; Fernández, F. M.; Grover, M. A.; Hud, N. V.; Schork, F. J.; Ferna, F. M.; Grover, M. A.; Hud, N. V.; Schork, F. J. *Macromolecules* **2014**, *47* (4), 1334–1343.

(127) Morávek, J. *Tetrahedron Lett.* **1967**, *8* (18), 1707–1710.

(128) Ogasawara, H.; Yoshida, A.; Imai, E. I.; Honda, H.; Hatori, K.; Matsuno, K. *Orig. Life Evol. Biosph.* **2000**, *30* (6), 519–526.

(129) Kawamura, K. *Int. J. Astrobiol.* **2005**, *3* (4), 301–309.

(130) Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39* (2), 99–123.

(131) LOHRMANN, R.; ORGEL, L. E. *Nature* **1973**, *244* (5416), 418–420.

(132) Khorana, H. G.; Jacob, T. M. *J. Am. Chem. Soc.* **1964**, *86* (8), 1630–1635.

(133) Sulston, J.; Lohrmann, R.; Orgel, L. E.; Miles, H. T. *Proc. Natl. Acad. Sci.* **1968**, *59* (3), 726–733.

(134) Sleeper, H. L.; Orgel, L. E. *J. Mol. Evol.* **1979**, *12* (4), 357–364.

(135) Monnard, P. A.; Kanavarioti, A.; Deamer, D. W. *J. Am. Chem. Soc.* **2003**, *125* (45), 13734–13740.

(136) Ferris, J. P. *Elements* **2005**, *1* (3), 145–149.

(137) Ferris, J. P.; Ertem, G. *J. Am. Chem. Soc.* **1993**, *115* (26), 12270–12275.

(138) Joshi, P. C.; Aldersley, M. F.; Delano, J. W.; Ferris, J. P. *J. Am. Chem. Soc.* **2009**, *131* (37), 13369–13374.

(139) Costanzo, G.; Saladino, R.; Crestini, C.; Ciciriello, F.; Di Mauro, E. *J. Biol. Chem.* **2007**, *282* (23), 16729–16735.

(140) Giovanna, C.; Pino, S.; Ciciriello, F.; Di Mauro, E. *J. Biol. Chem.* **2009**, *284* (48), 33206–33216.

(141) Costanzo, G.; Pino, S.; Botta, G.; Saladino, R.; Mauro, E. Di. *Orig. Life Evol. Biosph.* **2011**, *41* (6), 559–562.

(142) Morasch, M.; Mast, C. B.; Langer, J. K.; Schilcher, P.; Braun, D. *ChemBioChem* **2014**, *15* (6), 879–883.

(143) DeGuzman, V.; Vercoutere, W.; Shenasa, H.; Deamer, D. *J. Mol. Evol.* **2014**, *78* (5), 251–262.

(144) Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. *Orig. Life Evol. Biosph.* **2008**, *38* (1), 57–74.

(145) Da Silva, L.; Maurel, M.-C.; Deamer, D. *J. Mol. Evol.* **2015**, *80* (2), 86–97.

(146) Joyce, G F. Visser, GM. Van Voeckel, CA. Van Boom, JH. Orgel , LE. Van Westrenen, J. *Nature* **1984**, No. 310, 602–604.

(147) Blackmond, D. G. *Cold Spring Harbor perspectives in biology*. 2010, pp 117–125.

(148) Weissbuch, I.; Illos, R. A.; Bolbach, G.; Lahav, M. *Acc. Chem. Res.* **2009**, *42* (8), 1128–1140.

(149) Wieczorek, R.; Dörr, M.; Chotera, A.; Luisi, P. L.; Monnard, P.-A. *ChemBioChem* **2013**, *14* (2), 217–223.

(150) Yarus, M.; Widmann, J. J.; Knight, R. *J. Mol. Evol.* **2009**, *69* (5), 406–429.

(151) Turk-MacLeod, R. M.; Puthenvedu, D.; Majerfeld, I.; Yarus, M. *J. Mol. Evol.* **2012**, 1–9.

(152) Morasch, M.; Mast, C. B.; Langer, J. K.; Schilcher, P.; Braun, D. *ChemBioChem* **2014**, *15* (6), 879–883.

(153) Gorlero, M.; Wieczorek, R.; Adamala, K.; Giorgi, A.; Schininà, M. E.; Stano, P.; Luisi, P. L. *FEBS Lett.* **2009**, *583* (1), 153–156.

(154) Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R.-Y.; Algire, M. a; Benders, G. a; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. a; Young, L.; Qi, Z.-Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. a; Smith, H. O.; Venter, J. C. *Science* **2010**, *329* (5987), 52–56.

(155) Szostak, J. W. *J. Syst. Chem.* **2012**, *3* (1), 2.

(156) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409* (6818), 387–390.

(157) Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2008**, *37* (1), 101–108.

(158) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.; Sanders, J. K. M.; Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discov. Today* **2010**, *7* (2), 3652–3711.

(159) Li, J.; Nowak, P.; Otto, S. *J. Am. Chem. Soc.* **2013**, *135* (25), 9222–9239.

(160) Bissette, A. J.; Fletcher, S. P. *Angew. Chemie - Int. Ed.* **2013**, *52* (49), 12800–12826.

(161) Szostak, J. W. *Nature* **2009**, *459* (7244), 171–172.

(162) Weber, A. L. *Orig. Life Evol. Biosph.* **2001**, *31* (1/2), 71–86.

(163) Adamala, K.; Szostak, J. W. *Nat. Chem.* **2013**, *5* (6), 495–501.

(164) Gardner, P. M.; Winzer, K.; Davis, B. G. *Nat. Chem.* **2009**, *1* (5), 377–383.

(165) Deamer, D. W. *First Life: Discovering the connections between stars, planets, and evolution on earth.*; Berkeley: University of California Press., 2011.

(166) Gutierrez, J. M. P.; Hinkley, T.; Taylor, J. W.; Yanev, K.; Cronin, L. *Nat. Commun.* **2014**, *5*, 5571.

(167) Casadevall i Solvas, X.; DeMello, A. *Chem. Commun.* **2011**, *47* (7), 1936–1942.

(168) Song, H.; Chen, D. L.; Ismagilov, R. F. *Angew. Chem. Int. Ed. Engl.* **2006**, *45* (44), 7336–7356.

(169) Fradet, E.; McDougall, C.; Abbyad, P.; Dangla, R.; McGloin, D.; Baroud, C. N. *Lab Chip* **2011**, *11*, 4228.

(170) Vasiljevic, D.; Parojcic, J.; Primorac, M.; Vuleta, G. *Int. J. Pharm.* **2006**, *309* (1-2), 171–177.

(171) Deng, N.-N.; Meng, Z.-J.; Xie, R.; Ju, X.-J.; Mou, C.-L.; Wang, W.; Chu, L.-Y. *Lab Chip* **2011**, *11* (23), 3963–3969.

(172) Rojas, E. C.; Staton, J. A.; John, V. T.; Papadopoulos, K. D. *Langmuir* **2008**, *24* (14), 7154–7160.

(173) Kim, J.-W.; Utada, A. S.; Fernández-Nieves, A.; Hu, Z.; Weitz, D. A. *Angew. Chemie Int. Ed.* **2007**, *46* (11), 1819–1822.

(174) Xu, S.; Nie, Z.; Seo, M.; Lewis, P.; Kumacheva, E.; Stone, H. A.; Garstecki, P.; Weibel, D. B.; Gitlin, I.; Whitesides, G. M. *Angew. Chemie - Int. Ed.* **2005**, *44* (5), 724–728.

(175) Kim, S. H.; Abbaspourrad, A.; Weitz, D. A. *J. Am. Chem. Soc.* **2011**, *133* (14), 5516–5524.

(176) Chen, C. H.; Shah, R. K.; Abate, A. R.; Weitz, D. A. *Langmuir* **2009**, *25* (8), 4320–4323.

(177) Bernabini, C.; Holmes, D.; Morgan, H. *Lab Chip* **2011**, *11* (3), 407–412.

(178) Srinivasan, V.; Pamula, V. K.; Fair, R. B. *Lab Chip* **2004**, *4* (4), 310.

(179) Huebner, A.; Srisa-Art, M.; Holt, D.; Abell, C.; Hollfelder, F.; DeMello, A. J.; Edel, J. B. *Chem. Commun.* **2007**, *2* (12), 1218–1220.

(180) Brouzes, E.; Medkova, M.; Savenelli, N.; Marran, D.; Twardowski, M.; Hutchison, J. B.; Rothberg, J. M.; Link, D. R.; Perrimon, N.; Samuels, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (34), 14195–14200.

(181) Joensson, H. N.; Andersson Svahn, H. *Angew. Chemie - Int. Ed.* **2012**, *51* (49), 12176–12192.

(182) Song, H.; Tice, J. D.; Ismagilov, R. F. *Angew. Chemie - Int. Ed.* **2003**, *42* (7), 768–772.

(183) Zheng, B.; Roach, L. S.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2003**, *125* (37), 11170–11171.

(184) Song, H.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2003**, *125* (47), 14613–14619.

(185) Williams, R.; Peisajovich, S. G.; Miller, O. J.; Magdassi, S.; Tawfik, D. S.; Griffiths, A. D. *Nat. Methods* **2006**, *3* (7), 545–550.

(186) Kumaresan, P.; Yang, C. J.; Cronier, S. a; Blazej, R. G.; Mathies, R. a. *Anal. Chem.* **2008**, *80* (10), 3522–3529.

(187) Tawfik, D. S.; Griffiths, A. D. *Nat. Biotechnol.* **1998**, *16* (7), 652–656.

(188) Griffiths, A. D.; Tawfik, D. S. *Trends Biotechnol.* **2006**, *24* (9), 395–402.

(189) Kintses, B.; Hein, C.; Mohamed, M. F.; Fischlechner, M.; Courtois, F.; Lainé, C.; Hollfelder, F. *Chem. Biol.* **2012**, *19* (8), 1001–1009.

(190) Mazutis, L.; Baret, J.-C.; Treacy, P.; Skhiri, Y.; Araghi, A. F.; Ryckelynck, M.; Taly, V.; Griffiths, A. D. *Lab Chip* **2009**, *9* (20), 2902–2908.

(191) Baret, J.-C.; Miller, O. J.; Taly, V.; Ryckelynck, M.; El-Harrak, A.; Frenz, L.; Rick, C.; Samuels, M. L.; Hutchison, J. B.; Agresti, J. J.; Link, D. R.; Weitz, D. a; Griffiths, A. D. *Lab Chip* **2009**, *9* (13), 1850–1858.

(192) Mazutis, L.; Gilbert, J.; Ung, W. L.; Weitz, D. A.; Griffiths, A. D.; Heyman, J. A. *Nat. Protoc.* **2013**, *8* (5), 870–891.

(193) Zhao, M.; Nelson, W.; Wei, B. *Anal. …* **2013**, *20* (i).

(194) Teh, S.-Y.; Lin, R.; Hung, L.-H.; Lee, A. P. *Lab Chip* **2008**, *8* (2), 198–220.

(195) Theberge, A. B.; Courtois, F.; Schaerli, Y.; Fischlechner, M.; Abell, C.; Hollfelder, F.; Huck, W. T. S. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (34), 5846–5868.

(196) Love, J. C.; Anderson, J. R.; Whitesides, G. M. **2001**, No. July, 523–528.

(197) Whitesides, G. M. *Nature* **2006**, *442* (7101), 368–373.

(198) Martino, C.; DeMello, A. J. *Interface Focus* **2016**, *6* (4), 20160011.

(199) Mann, S. *Angew. Chemie Int. Ed.* **2013**, *52* (1), 155–162.

(200) Miras, H. N.; Cooper, G. J. T.; Long, D.-L.; Bogge, H.; Muller, A.; Streb, C.; Cronin, L. *Science (80-.).* **2010**, *327* (5961), 72–74.

(201) Cooper, G. J. T.; Bowman, R. W.; Magennis, E. P.; Fernandez-Trillo, F.; Alexander, C.; Padgett, M. J.; Cronin, L. *Angew. Chemie Int. Ed.* **2012**, *51* (51), 12754–12758.

(202) GANTI, T. *Cell Biol. Int.* **2002**, *26* (8), 729–735.

(203) Gayon, J.; Malaterre, C.; Morange, M.; Raulin-Cerceau, F.; Tirard, S. *Orig. Life Evol. Biosph.* **2010**, *40* (2), 119–120.

(204) Mata, A.; Fleischman, A. J. **2005**, *2*, 281–293.

(205) Esteves, A. C. C.; Brokken-Zijp, J.; Laven, J.; Huinink, H. P.; Reuvers, N. J. W.; Van, M. P.; de With, G. *Polymer (Guildf).* **2009**, *50* (16), 3955–3966.

(206) Abate, A. R.; Thiele, J.; Weinhart, M.; Weitz, D. a. *Lab Chip* **2010**, *10* (14), 1774–1776.

(207) Baret, J.-C. *Lab Chip* **2012**, *12* (3), 422–433.

(208) Theberge, A. B.; Mayot, E.; El Harrak, A.; Kleinschmidt, F.; Huck, W. T. S.; Griffiths, A. D. *Lab Chip* **2012**, *12* (7), 1320–1326.

(209) Abbyad, P.; Dangla, R.; Alexandrou, A.; Baroud, C. N. *Lab Chip* **2011**, *11* (5), 813–821.

(210) Agresti, J. J.; Antipov, E.; Abate, A. R.; Ahn, K.; Rowat, A. C.; Baret, J.-C.; Marquez, M.; Klibanov, A. M.; Griffiths, A. D.; Weitz, D. a. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (9), 4004–4009.

(211) Abate, A. R.; Agresti, J. J.; Weitz, D. a. *Appl. Phys. Lett.* **2010**, *96* (20), 1–3.

(212) Paegel, B. M.; Joyce, G. F. *Chem. Biol.* **2010**, *17* (7), 717–724.

(213) Paegel, B. M.; Joyce, G. F. *PLoS Biol.* **2008**, *6* (4), e85.

(214) Fallah-Araghi, A.; Meguellati, K.; Baret, J. C.; El Harrak, A.; Mangeat, T.; Karplus, M.; Ladame, S.; Marques, C. M.; Griffiths, A. D. *Phys. Rev. Lett.* **2014**, *112* (2), 1–5.

(215) Shiomi, H.; Tsuda, S.; Suzuki, H.; Yomo, T. *PLoS One* **2014**, *9* (7), 5–10.

(216) Lahav, N.; White, D.; Chang, S. *Science* **1978**, *201* (JULY 1978), 67–69.

(217) Shanker, U.; Bhushan, B.; Bhattacharjee, G.; Kamaluddin. *Orig. Life Evol. Biosph.* **2012**, *42* (1), 31–45.

(218) Moreau, P.; Dehmoune, J.; Salmon, J. B.; Leng, J. *Appl. Phys. Lett.* **2009**, *95* (3).

(219) Bujd**?**k, J.; Son, H.; Yongyai, Y.; Rode, B. M. *Catal. Letters* **1996**, *37* (3-4), 267–272.

(220) Fuchida, S.; Masuda, H.; Shinoda, K. *Orig. Life Evol. Biosph.* **2014**, *44* (1), 13–28.

(221) Fitz, D.; Reiner, H.; Rode, B. M. *Pure Appl. Chem.* **2007**, *79* (12), 2101–2117.

(222) Perry, R. *Perry's Chemical Engineers' Handbook (6th ed.)*; McGraw Hill, 1997.

(223) Viedma, C. **2000**, 549–556.

(224) Clayden, J.; Greeves, N.; Warren, S. *Organic Chemistry*; OUP Oxford, 2012.

(225) Rufo, C. M.; Moroz, Y. S.; Moroz, O. V; Stöhr, J.; Smith, T. a; Hu, X.; DeGrado, W. F.; Korendovych, I. V. *Nat. Chem.* **2014**, *6* (4), 303–309.

(226) Murray, K. K.; Boyd, R. K.; Eberlin, M. N.; Langley, G. J.; Li, L.; Naito, Y. *Pure Appl. Chem* **2013**, *85* (7), 1515–1609.

(227) Longo, L. M.; Lee, J.; Blaber, M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (6), 2135–2139.

(228) Meierhenrich, U. *Amino Acids and the Asymmetry of Life*; 2008; Vol. 14.

(229) Bélières, M.; Chouini-Lalanne, N.; Déjugnat, C. *RSC Adv.* **2015**, *5* (45), 35830–35842.

(230) Kitadai, N.; Yokoyama, T.; Nakashima, S. *Geochim. Cosmochim. Acta* **2011**, *75* (21), 6285–6299.

(231) Kitadai, N.; Oonishi, H.; Umemoto, K.; Usui, T.; Fukushi, K.; Nakashima, S. *Orig. Life Evol. Biosph.* **2016**.

(232) Vauthey, S.; Santoso, S.; Gong, H.; Watson, N.; Zhang, S. *Proc. Natl. Acad. Sci.* **2002**, *99* (8), 5355–5360.

(233) Kamat, N. P.; Tob??, S.; Hill, I. T.; Szostak, J. W. *Angew. Chemie - Int. Ed.* **2015**, *54* (40), 11735–11739.

(234) Jia, T. Z.; Fahrenbach, A. C.; Kamat, N. P.; Adamala, K. P.; Szostak, J. W. *Nat. Chem.* **2016**, No. June, In Press.

(235) Forsythe, J. G.; Yu, S.-S. S.; Mamajanov, I.; Grover, M. A.; Krishnamurthy, R.; Fernandez, F. M.; Hud, N. V; Fern??ndez, F. M.; Hud, N. V. *Angew Chem Int Ed Engl* **2015**, *10*, n/a???n/a.

(236) Deamer, D. *Chem. Soc. Rev.* **2012**, *41* (16), 5375–5379.

(237) Mungi, C.; Rajamani, S. *Life* **2015**, *5* (1), 65–84.

(238) Team, R. D. C.; R Development Core Team, R. *R Found. Stat. Comput.* **2005**, *1* (2.11.1), 409.

(239) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. a; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. *Nat. Biotechnol.* **2012**, *30* (10), 918–920.

(240) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78* (3), 779–787.

7 Appendix

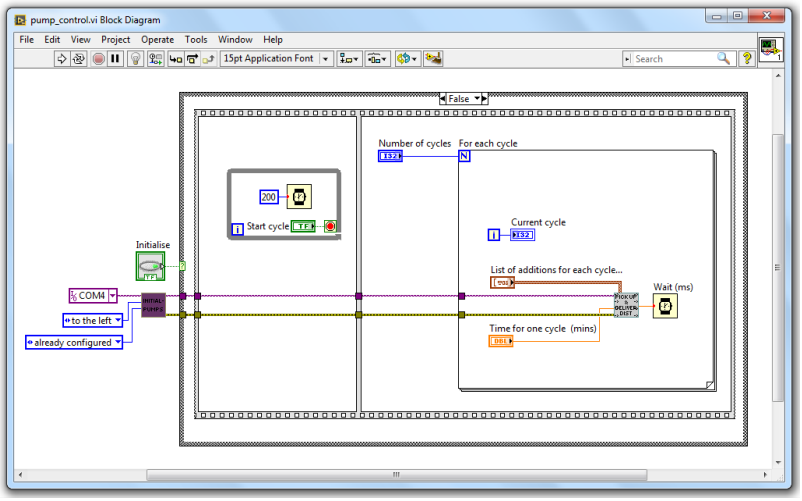


Figure A1: The National Instruments LabVIEW software Block Diagram associated with the Front Panel in Fig. A2. The pump configuration is hard-wired into an initialisation Sub VI. Shown is the non-initialisation case, the software waits for the user to click "Start Cycle". A "For" loop, named "For each cycle," then runs for the "Number of Cycles," set up the user. A Sub VI named "Pick up and Deliver (distribution vale).vi," reads the "List of additions for each cycle…," and executes them in parallel, then outputs the remaining "Wait (ms)" time in milliseconds which is used to delay to next iteration of the experiment.

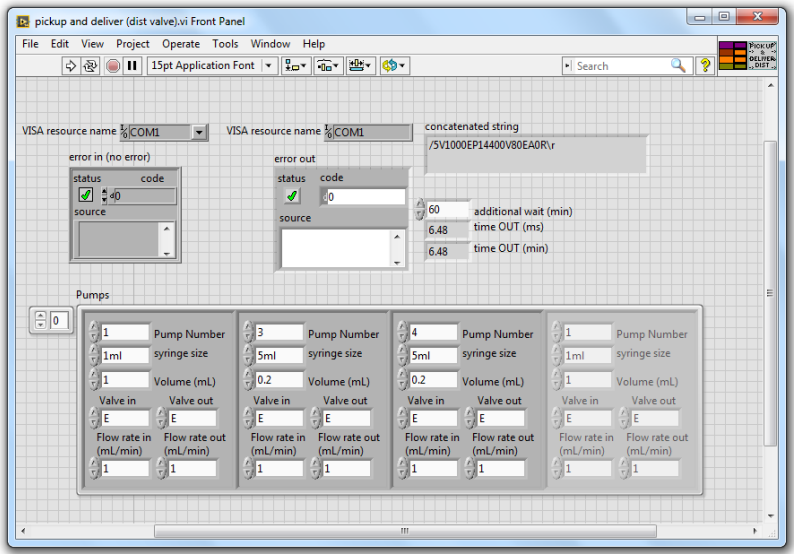


Figure A2: A screen shot of the front panel of the "Pick up and Deliver (distribution valve).vi" Sub VI as used by the "pump_control.vi". The function of this VI is to convert the user defined instructions in the "Pumps" array into a string command (named "concatenated string") which is sent via serial to the each pump and valve unit.

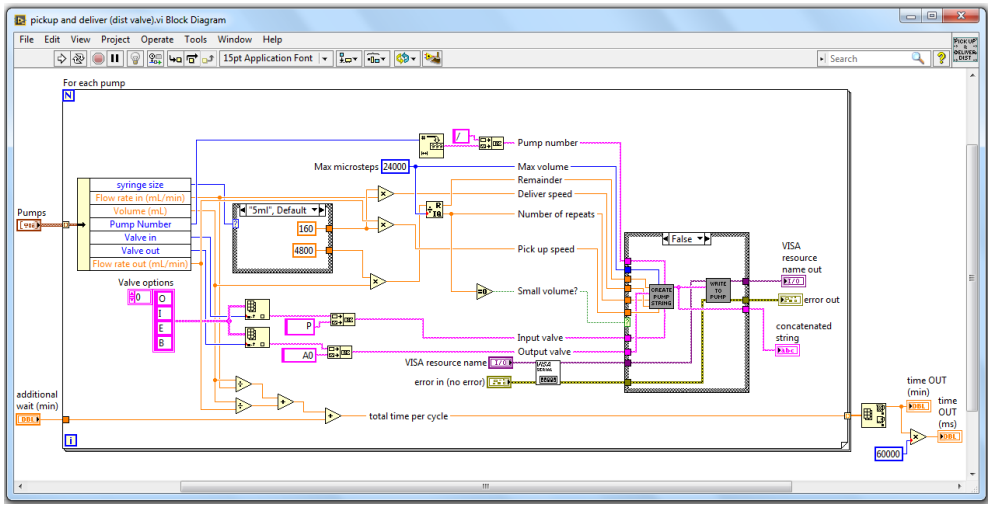


Figure A3: The Block Diagram associated with the Front Panel in Fig. A2. This VI converts the user instructions in the "Pumps" array (equal to "List of additions for each cycle…") into a string command, and then sends that command to the correct pump. This is done in a "For" loop, named "For each pump," whereby each pump is addressed sequentially (the order is determined by the order they appear in the "Pumps" array). The VI also calculates the time to wait between each iteration of the experiment, given by the sum of the time taken to operate the pumps, plus the user defined "additional wait (min)".

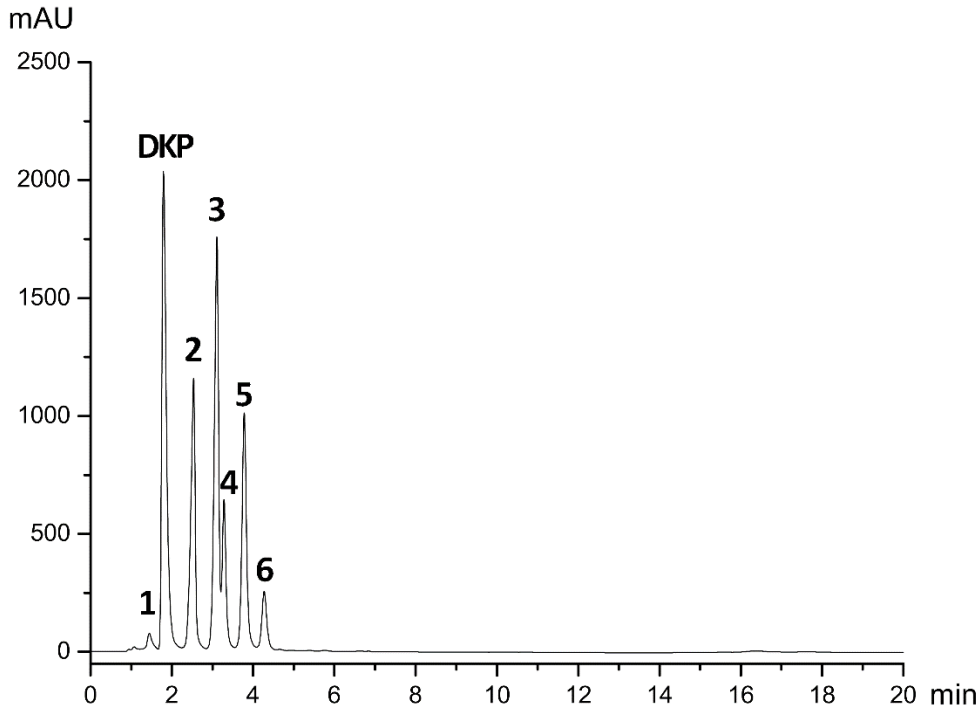


Figure A4: IP-HPLC chromatogram of Glyn oligomer standards.

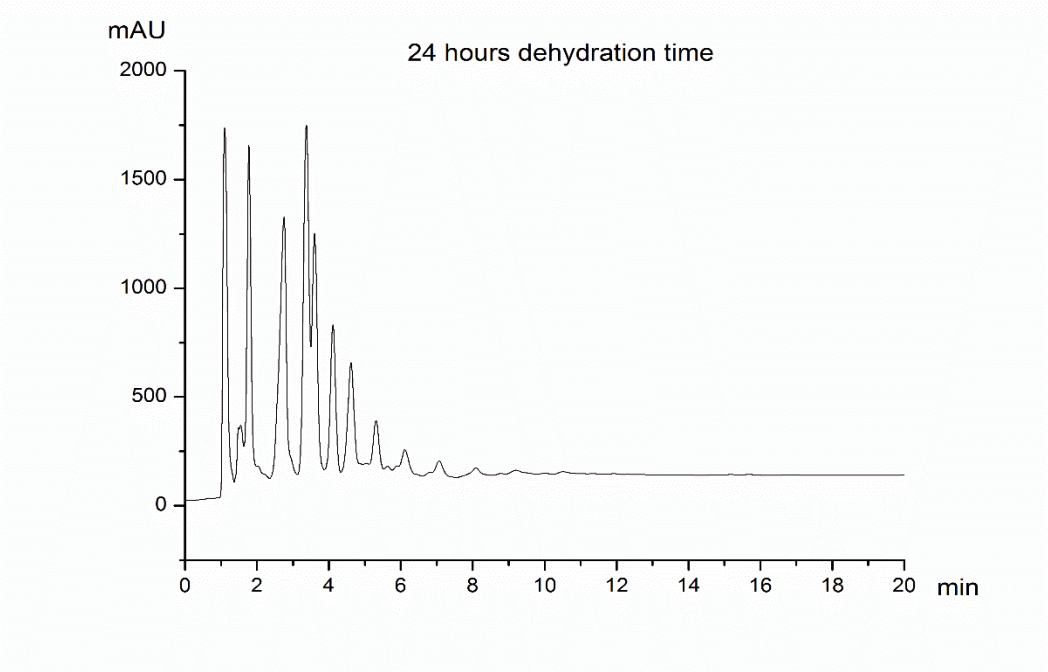


Figure A5: IP-HPLC chromatogram of a standard glycine oligomerisation reaction after 24h at 130ºC with initial pH=9.8.

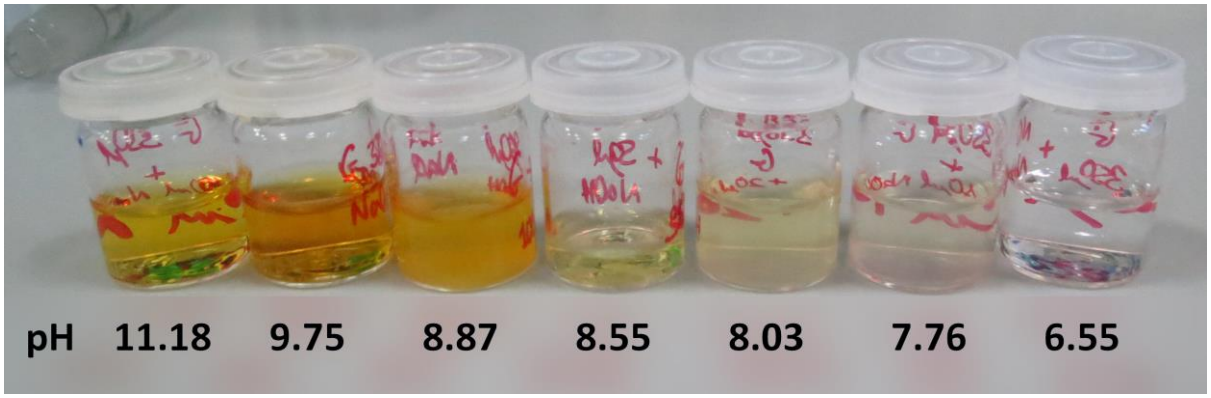


Figure A6: Correlation between the pH and the browning produced by decomposition processes.

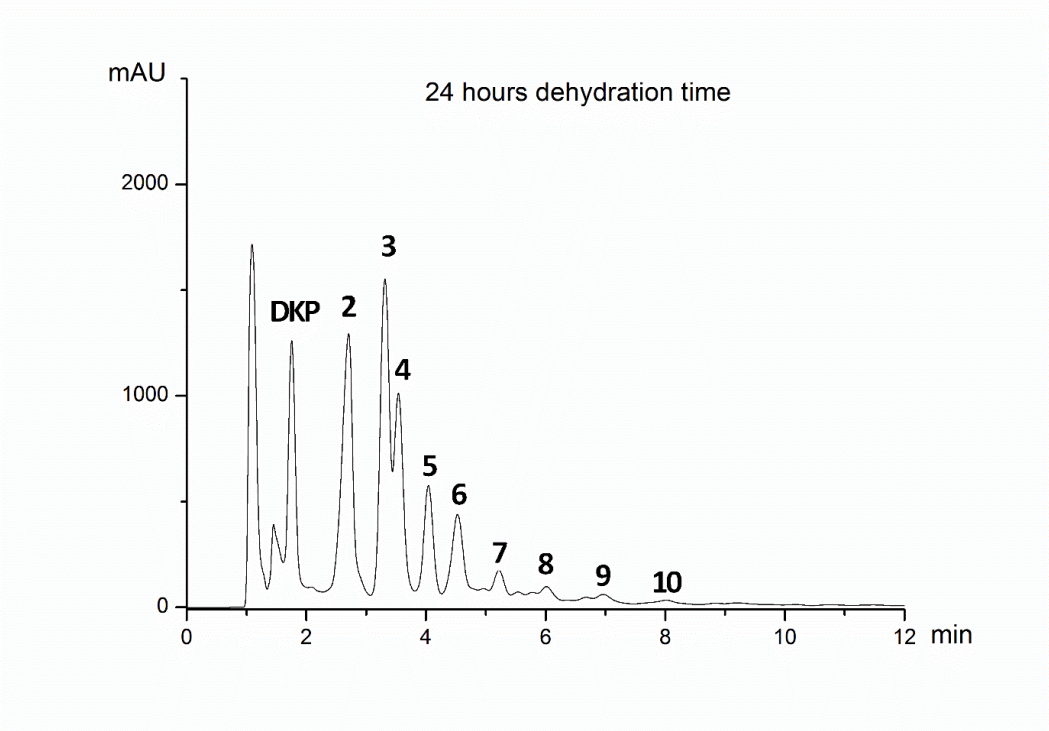


Figure A7: IP-HPLC chromatogram of glycine oligomerisation products (24h, 130ºC) where a Teflon reaction vessel was used.

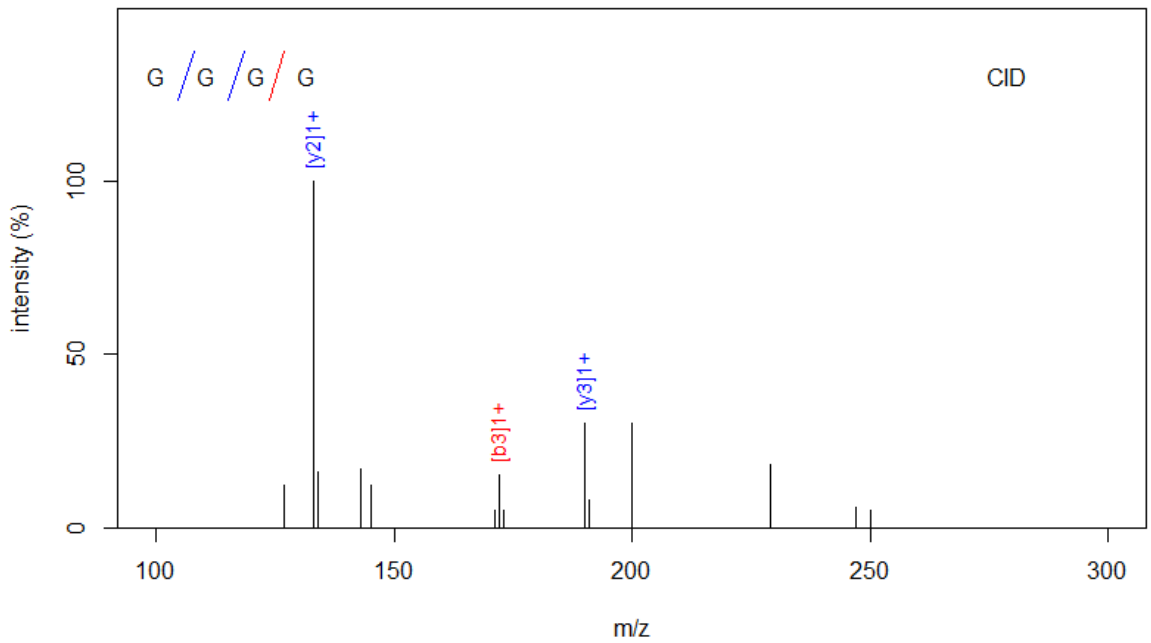


Figure A8: MS/MS of fragmentation products of ion assigned as Gly4 oligomer (m/z = 247.104); experimental data matches fragmentation products calculated for GGGG sequence.

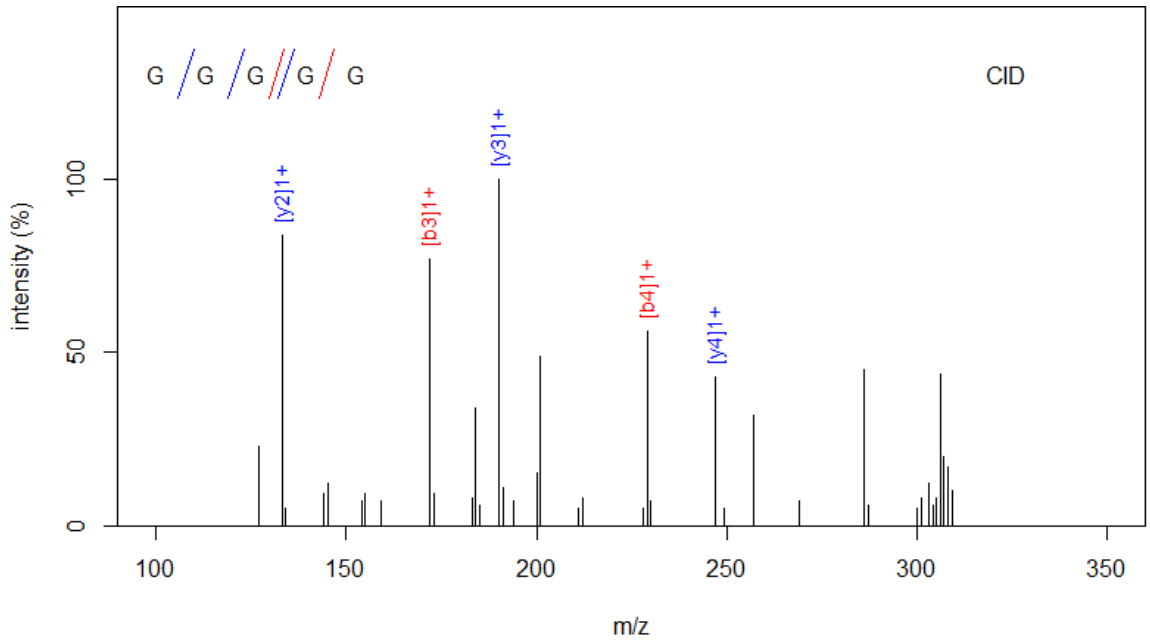


Figure A9: MS/MS of fragmentation products of ion assigned as Gly5 oligomer (m/z = 304.125); experimental data matches fragmentation products calculated for GGGGG sequence.

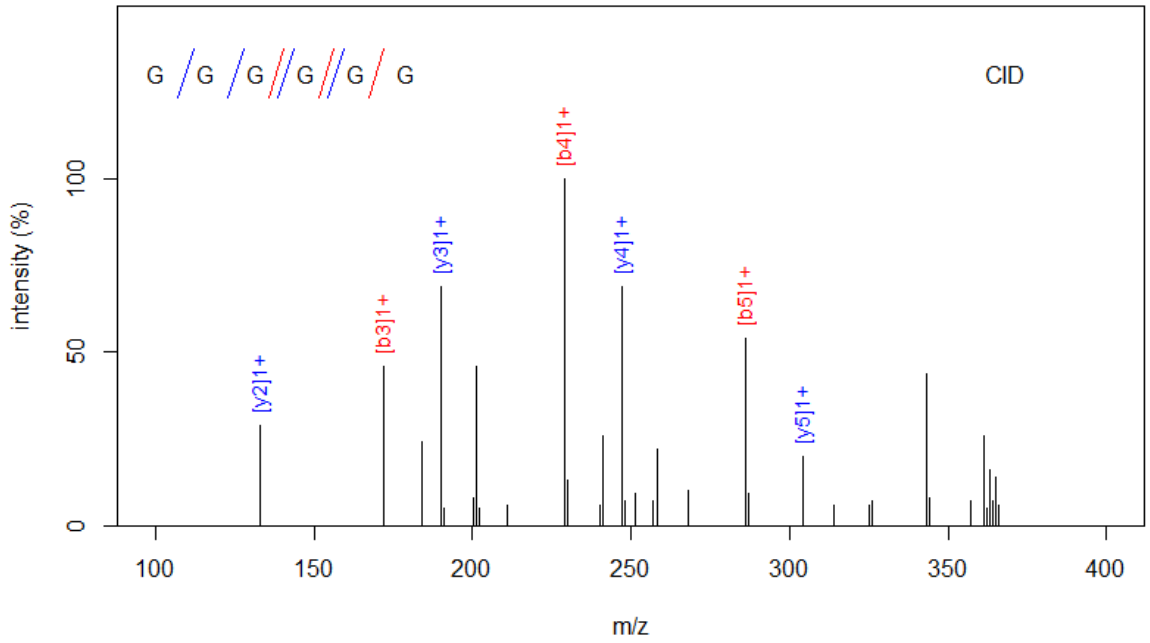


Figure A10: MS/MS of fragmentation products of ion assigned as Gly6 oligomer (m/z = 361.147); experimental data matches fragmentation products calculated for GGGGGG sequence.

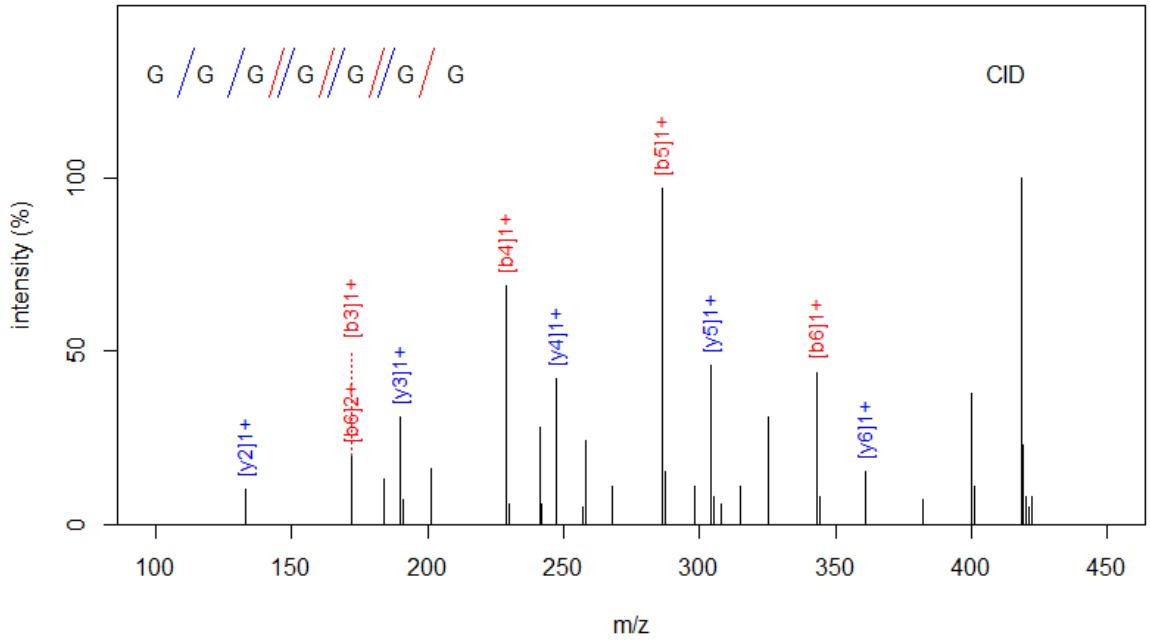


Figure A11: MS/MS of fragmentation products of ion assigned as Gly7 oligomer (m/z = 418.168); experimental data matches fragmentation products calculated for GGGGGGG sequence.

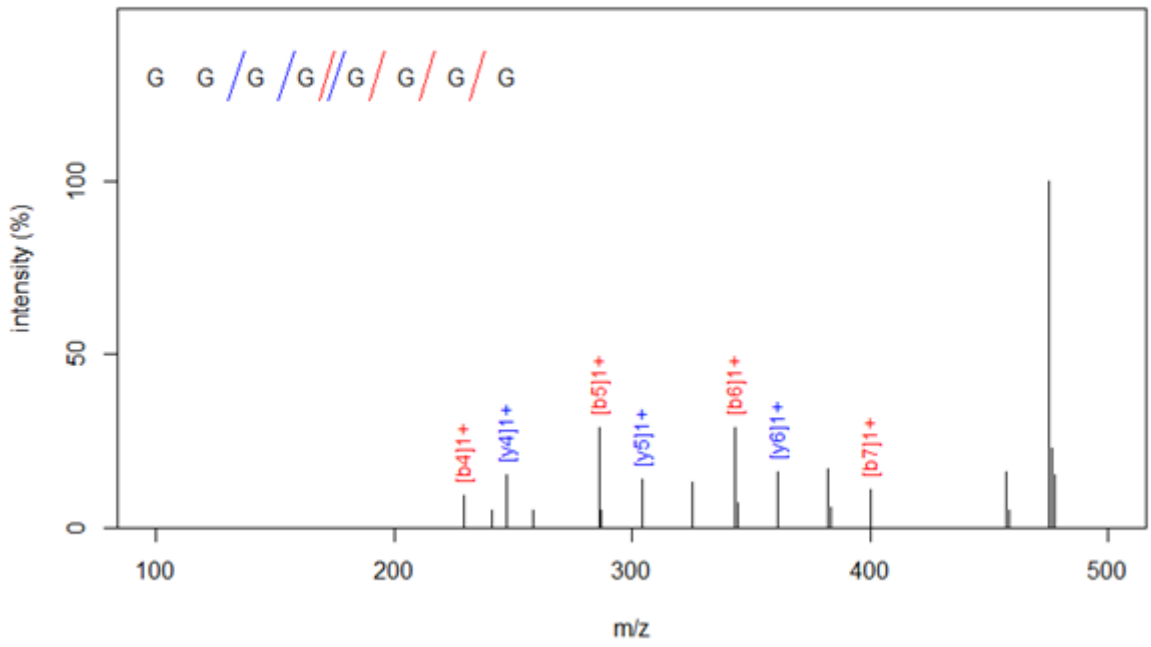


Figure A12: MS/MS of fragmentation products of ion assigned as Gly8 oligomer (m/z = 475.190); experimental data matches fragmentation products calculated for GGGGGGGGG sequence.

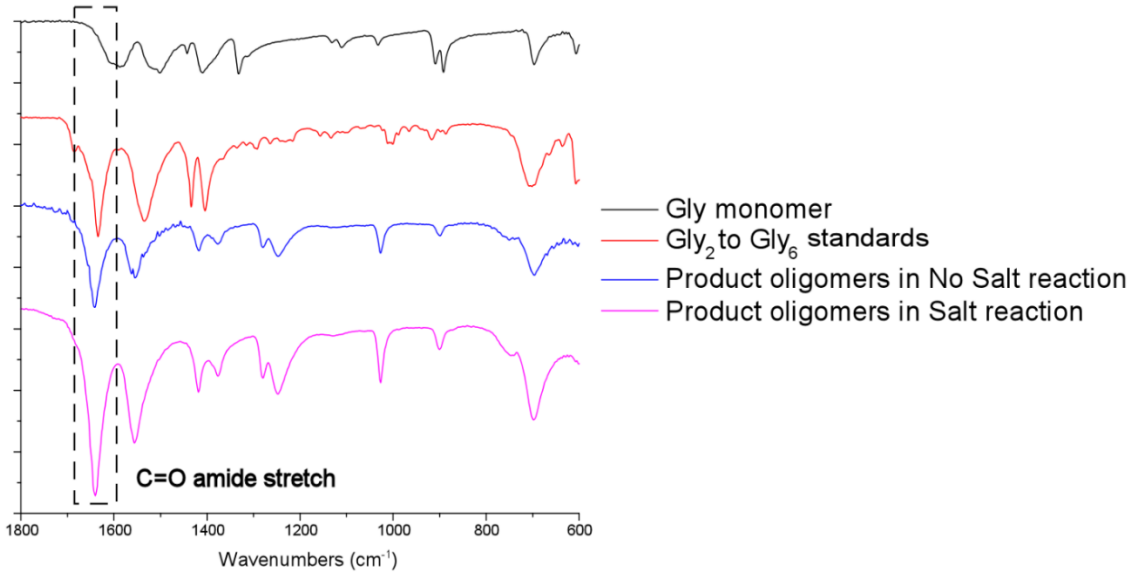


Figure A13: C=O amide stretching IR band of glycine monomer, gly2-gly6 standard oligomers, and two product oligomer samples run with different NaCl concentrations. As can be seen the IR of our samples indeed matches the oligomers which confirms our HPLC results.

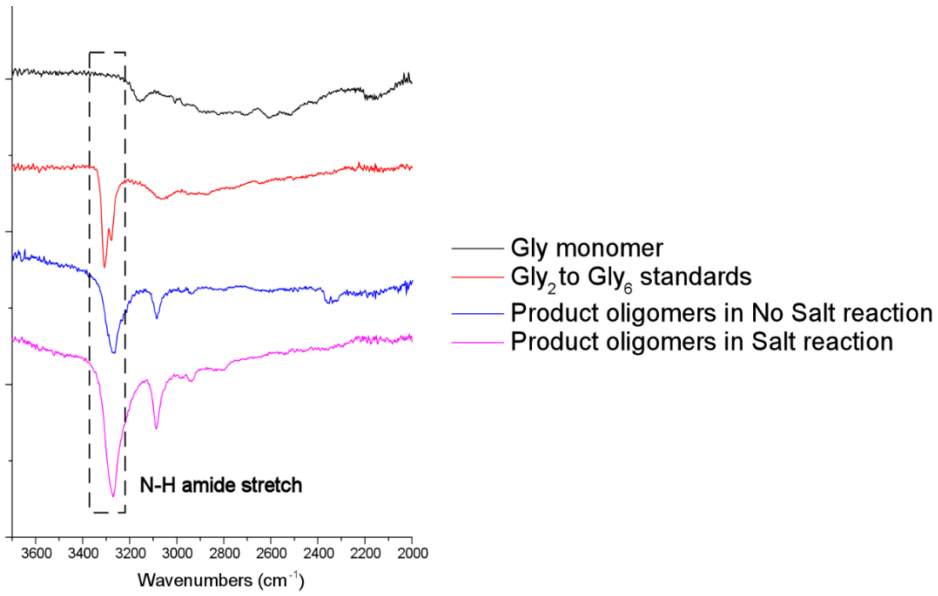


Figure A14: N-H amide stretching IR band of glycine monomer, gly2-gly6 standard oligomers, and two product oligomer samples ran with different NaCl concentrations. As can be seen the IR of our samples indeed matches the oligomers which confirms our HPLC results.

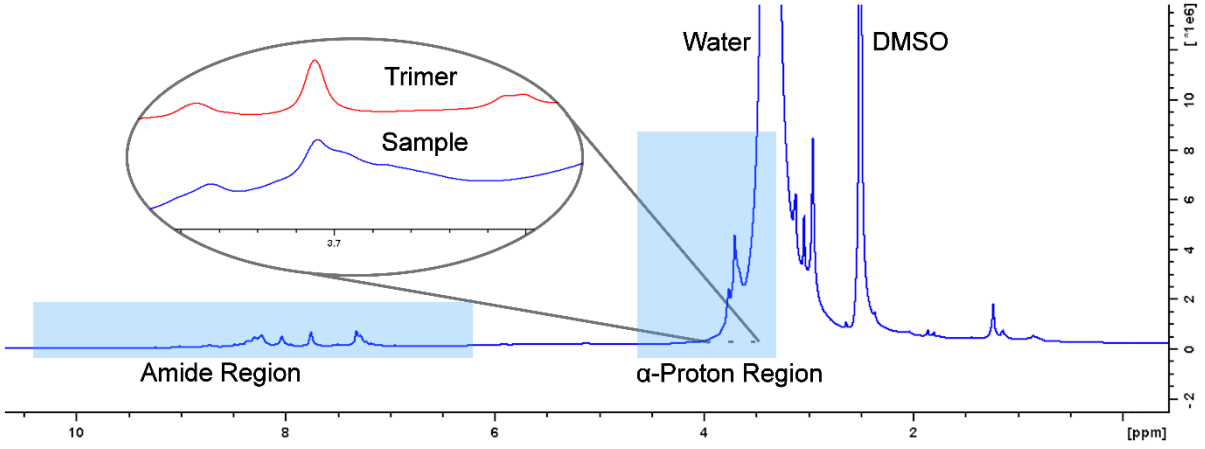


Figure A15: 1H-NMR spectrum of APS products in DMSO highlighting amide and α-proton regions.

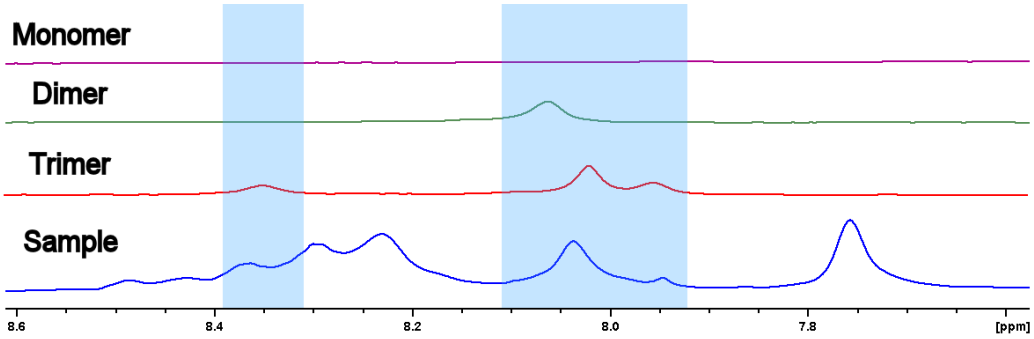


Figure A16: 1H-NMR comparison of Gly, Gly2, Gly3 and the APS sample in the amide region.

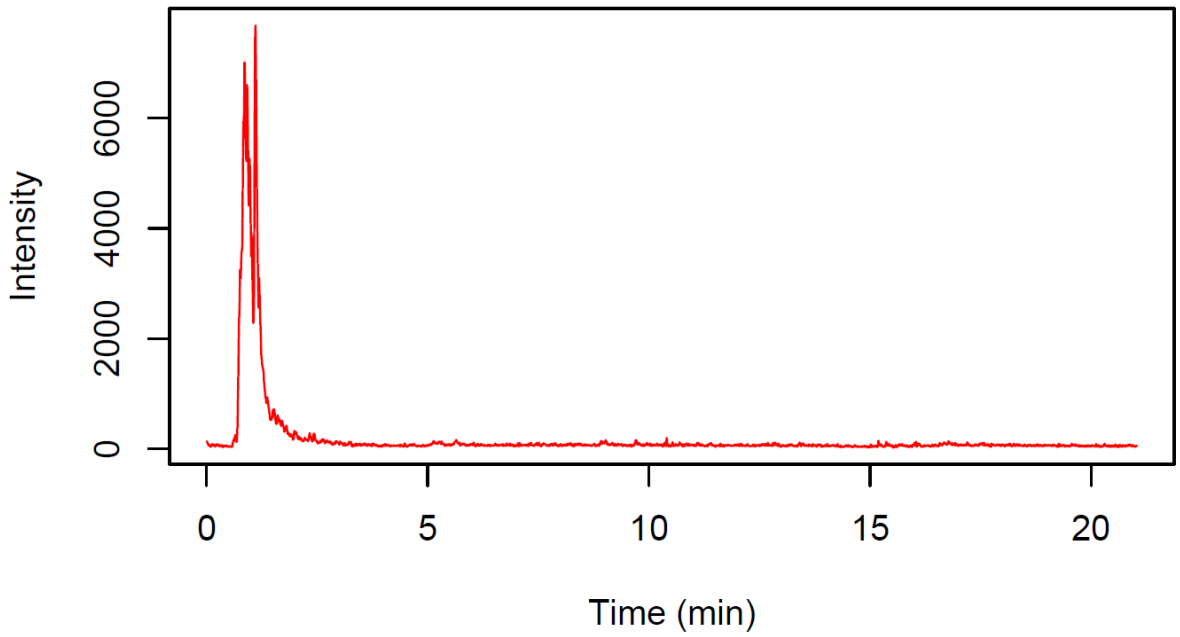


Figure A17: "Virtual" BPC traces illustrating co-condensation of G and A monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
90.054	GGAA	A	[y1]1+	90.055	-0.001
90.054	GAGA	A	[y1]1+	90.055	-0.001
90.054	AGGA	A	[y1]1+	90.055	-0.001
115.05	GGAA	GG	[b2]1+	115.05	0
129.0656	GAGA	GA	[b2]1+	129.066	0
129.0656	GAAG	GA	[b2]1+	129.066	0
129.0656	AGGA	AG	[b2]1+	129.066	0
129.0656	AGAG	AG	[b2]1+	129.066	0
133.0613	AAGG	GG	[y2]1+	133.061	0
147.0752	GAGA	GA	[y2]1+	147.076	-0.001
147.0752	GAAG	AG	[y2]1+	147.076	-0.001
147.0752	AGGA	GA	[y2]1+	147.076	-0.001
147.0752	AGAG	AG	[y2]1+	147.076	-0.001
186.0859	GGAA	GGA	[b3]1+	186.087	-0.001
186.0859	GAGA	GAG	[b3]1+	186.087	-0.001
186.0859	AGGA	AGG	[b3]1+	186.087	-0.001
204.0965	AGGA	GGA	[y3]1+	204.098	-0.002
204.0965	AGAG	GAG	[y3]1+	204.098	-0.002
204.0965	AAGG	AGG	[y3]1+	204.098	-0.002

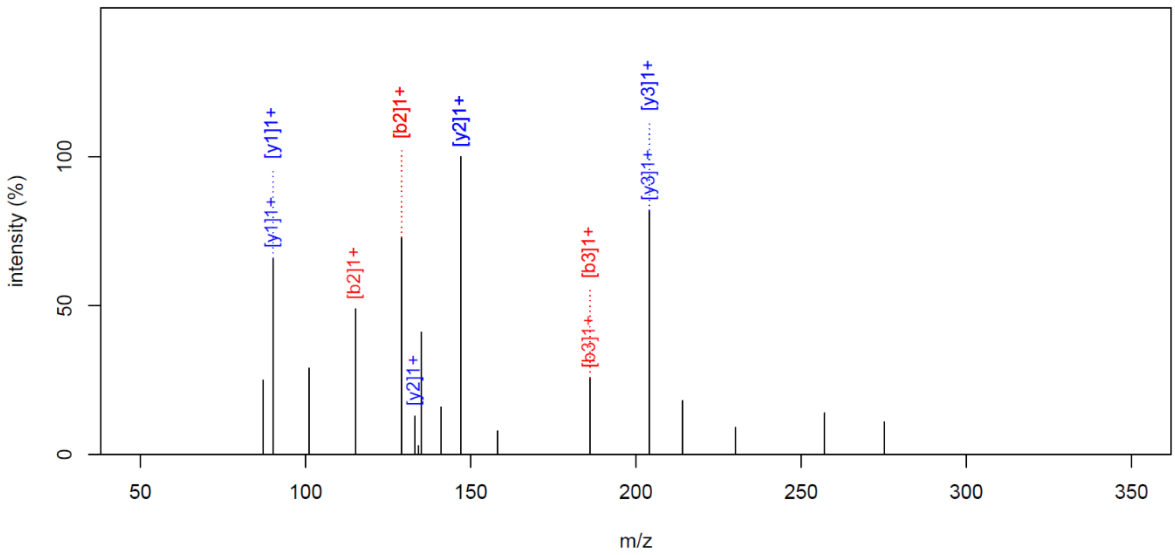


Figure A18: Example MS/MS analyses of products of G and A monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 275.135 (corresponding to a tetramer incorporating 2 x G and 2x A). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x A (ie. more than one sequence of the same composition is present, all incorporating G and A).

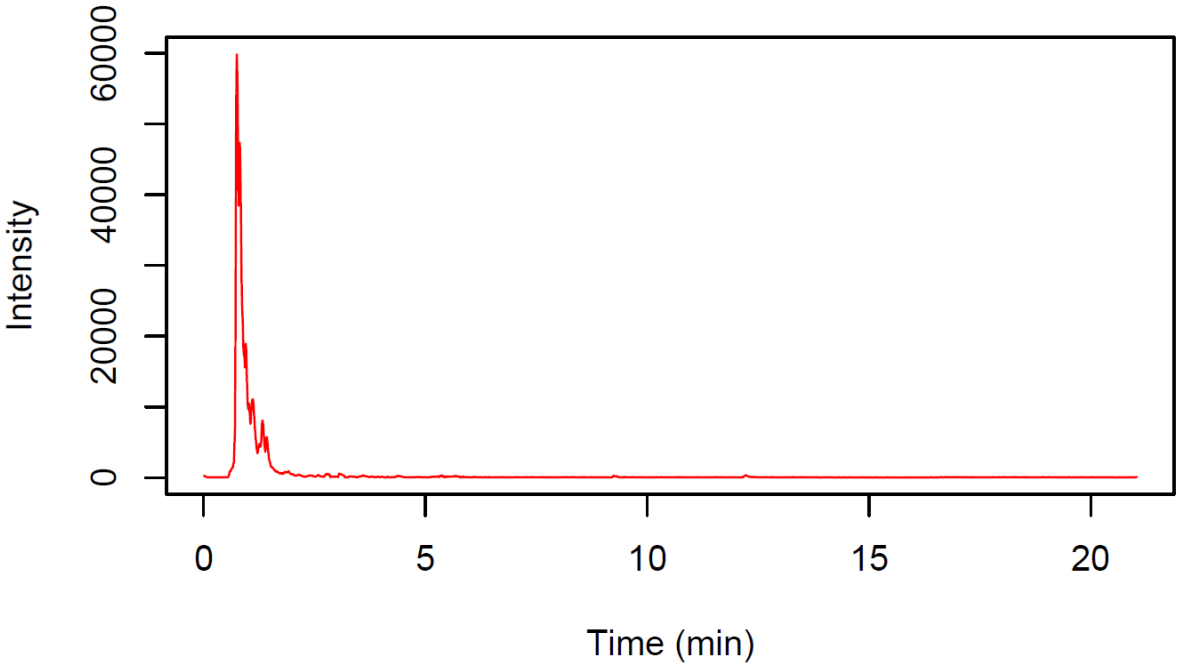


Figure A19: “Virtual” BPC traces illustrating co-condensation of G and D monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
76.0327	GDDG	G	[y1]1+	76.039	-0.006
76.0327	DGDG	G	[y1]1+	76.039	-0.006
76.0327	DDGG	G	[y1]1+	76.039	-0.006
115.051	GGDD	GG	[b2]1+	115.05	0.001
173.0516	GDGD	GD	[b2]1+	173.056	-0.004
173.0516	GDDG	GD	[b2]1+	173.056	-0.004
173.0516	DGGD	DG	[b2]1+	173.056	-0.004
173.0516	DGDG	DG	[b2]1+	173.056	-0.004
191.0673	GDGD	GD	[y2]1+	191.066	0.001
191.0673	GDDG	DG	[y2]1+	191.066	0.001
191.0673	DGGD	GD	[y2]1+	191.066	0.001
191.0673	DGDG	DG	[y2]1+	191.066	0.001
230.0741	GGDD	GGD	[b3]1+	230.077	-0.003
230.0741	GDGD	GDG	[b3]1+	230.077	-0.003
230.0741	DGGD	DGG	[b3]1+	230.077	-0.003
231.0617	DDGG	DD	[b2]1+	231.061	0.001
248.0845	DGGD	GGD	[y3]1+	248.088	-0.004
248.0845	DGDG	GDG	[y3]1+	248.088	-0.004
248.0845	DDGG	DGG	[y3]1+	248.088	-0.004
249.0646	GGDD	DD	[y2]1+	249.072	-0.007
288.077	GDDG	GDD	[b3]1+	288.083	-0.006
288.077	DGDG	DGD	[b3]1+	288.083	-0.006
288.077	DDGG	DDG	[b3]1+	288.083	-0.006
306.091	GGDD	GDD	[y3]1+	306.093	-0.002
306.091	GDGD	DGD	[y3]1+	306.093	-0.002
306.091	GDDG	DDG	[y3]1+	306.093	-0.002

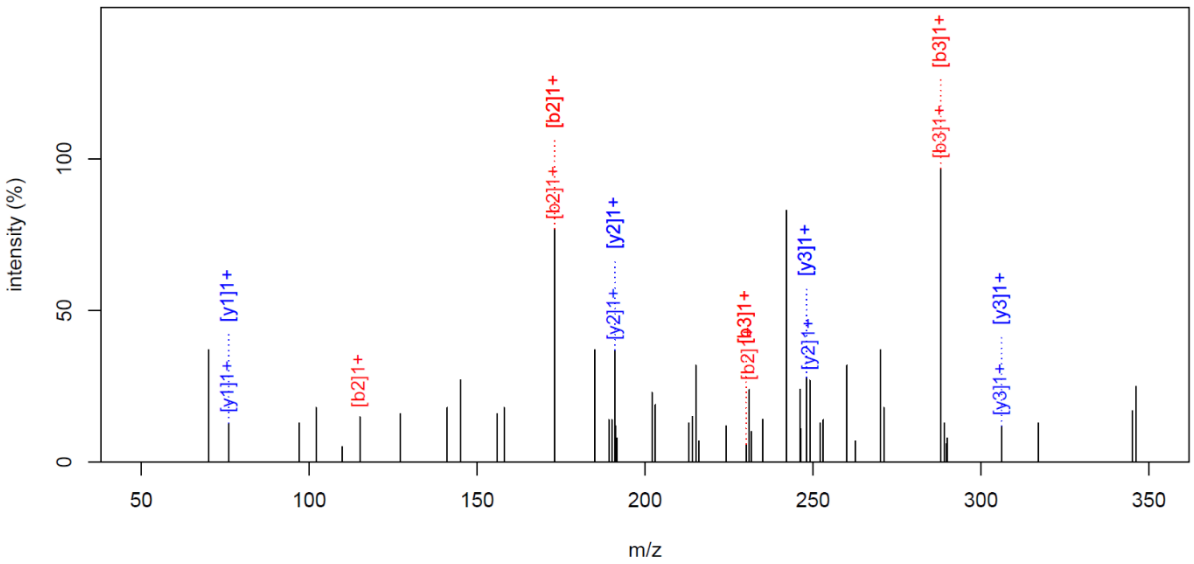


Figure A20: Example MS/MS analyses of products of G and D monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 363.115 (corresponding to a tetramer incorporating 2 x G and 2 x D). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x D (ie. more than one sequence of the same composition is present, all incorporating G and D).

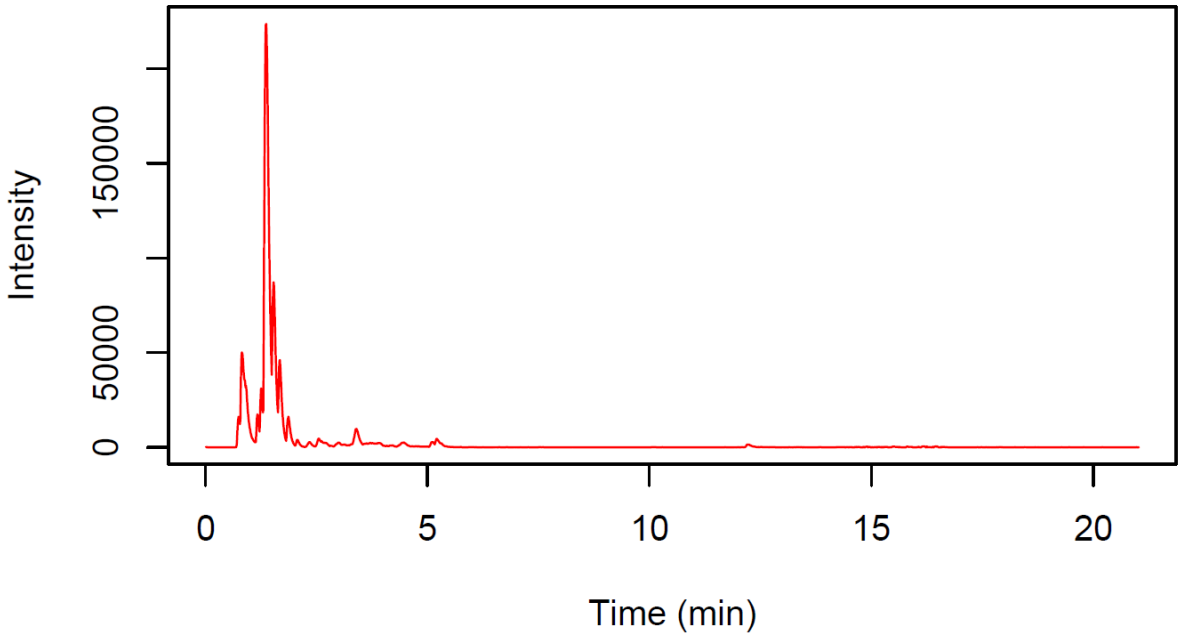


Figure A21: "Virtual" BPC traces illustrating co-condensation of G and E monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
130.0536	EGGE	E	[b1]1+	130.05	0.004
130.0536	EGEG	E	[b1]1+	130.05	0.004
130.0536	EEGG	E	[b1]1+	130.05	0.004
187.0693	GEGE	GE	[b2]1+	187.071	-0.002
187.0693	GEEG	GE	[b2]1+	187.071	-0.002
187.0693	EGGE	EG	[b2]1+	187.071	-0.002
187.0693	EGEG	EG	[b2]1+	187.071	-0.002
205.0784	GEGE	GE	[y2]1+	205.082	-0.004
205.0784	GEEG	EG	[y2]1+	205.082	-0.004
205.0784	EGGE	GE	[y2]1+	205.082	-0.004
205.0784	EGEG	EG	[y2]1+	205.082	-0.004
244.1021	GGEE	GGE	[b3]1+	244.093	0.009
244.1021	GEGE	GEG	[b3]1+	244.093	0.009
244.1021	EGGE	EGG	[b3]1+	244.093	0.009
259.0765	EEGG	EE	[b2]1+	259.092	-0.015
262.108	EGGE	GGE	[y3]1+	262.103	0.005
262.108	EGEG	GEG	[y3]1+	262.103	0.005
262.108	EEGG	EGG	[y3]1+	262.103	0.005
316.0786	GEEG	GEE	[b3]1+	316.114	-0.035
316.0786	EGEG	EGE	[b3]1+	316.114	-0.035
316.0786	EEGG	EEG	[b3]1+	316.114	-0.035

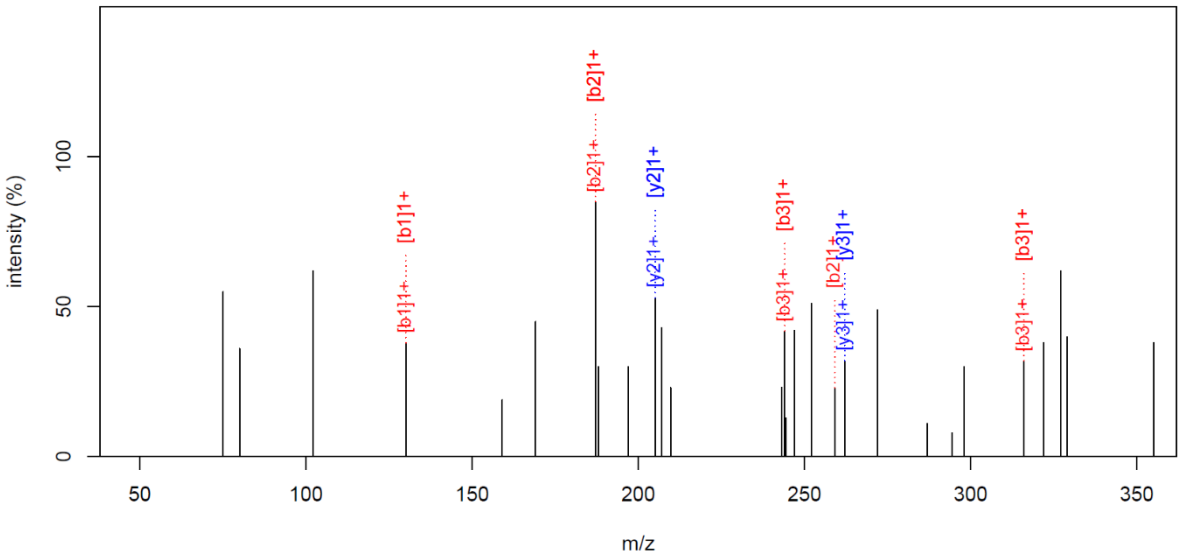


Figure A22: Example MS/MS analyses of products of G and E monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 391.146 (corresponding to a tetramer incorporating 2 x G and 2 x E). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x E (ie. more than one sequence of the same composition is present, all incorporating G and E).

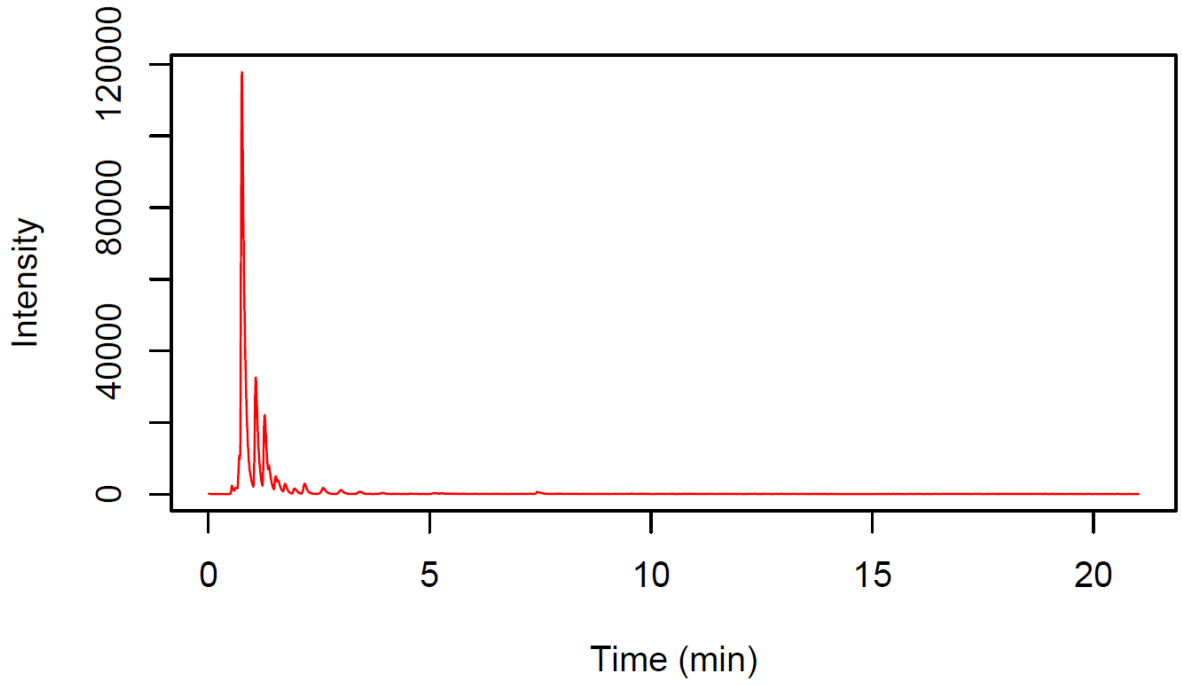


Figure A23: “Virtual” BPC traces illustrating co-condensation of G and K monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
129.0966	KKGG	K	[b1]1+	129.102	-0.005
129.0966	KGKG	K	[b1]1+	129.102	-0.005
129.0966	KGGK	K	[b1]1+	129.102	-0.005
147.1163	KGGK	K	[y1]1+	147.113	0.003
147.1163	GKGK	K	[y1]1+	147.113	0.003
147.1163	GGKK	K	[y1]1+	147.113	0.003
186.105	KGKG	KG	[b2]1+	186.124	-0.019
186.105	KGGK	KG	[b2]1+	186.124	-0.019
186.105	GKKG	GK	[b2]1+	186.124	-0.019
186.105	GKGK	GK	[b2]1+	186.124	-0.019
204.1326	KGKG	KG	[y2]1+	204.134	-0.001
204.1326	KGGK	GK	[y2]1+	204.134	-0.001
204.1326	GKKG	KG	[y2]1+	204.134	-0.001
204.1326	GKGK	GK	[y2]1+	204.134	-0.001
243.1491	KGGK	KGG	[b3]1+	243.145	0.004
243.1491	GKGK	GKG	[b3]1+	243.145	0.004
243.1491	GGKK	GGK	[b3]1+	243.145	0.004
257.1957	KKGG	KK	[b2]1+	257.197	-0.001
261.1581	KKGG	KGG	[y3]1+	261.156	0.002
261.1581	KGKG	GKG	[y3]1+	261.156	0.002
261.1581	KGGK	GGK	[y3]1+	261.156	0.002
275.2036	GGKK	KK	[y2]1+	275.208	-0.004
332.2238	GKKG	KKG	[y3]1+	332.229	-0.005
332.2238	GKGK	KGK	[y3]1+	332.229	-0.005
332.2238	GGKK	GKK	[y3]1+	332.229	-0.005

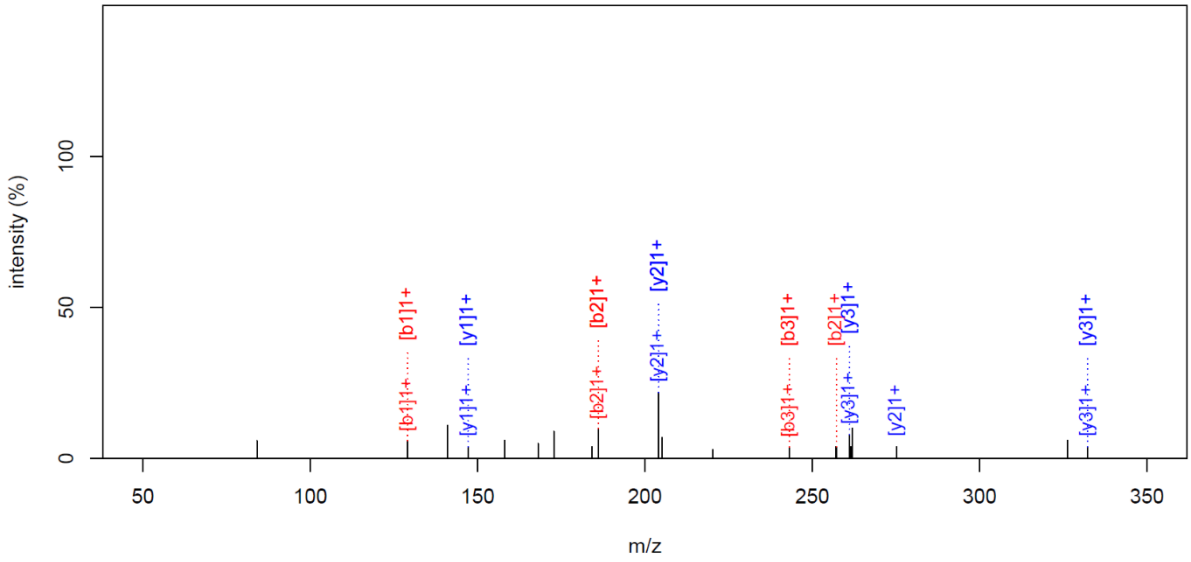


Figure A24: Example MS/MS analyses of products of G and K monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 389.251 (corresponding to a tetramer incorporating 2 x G and 2 x K). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x K (ie. more than one sequence of the same composition is present, all incorporating G and K).

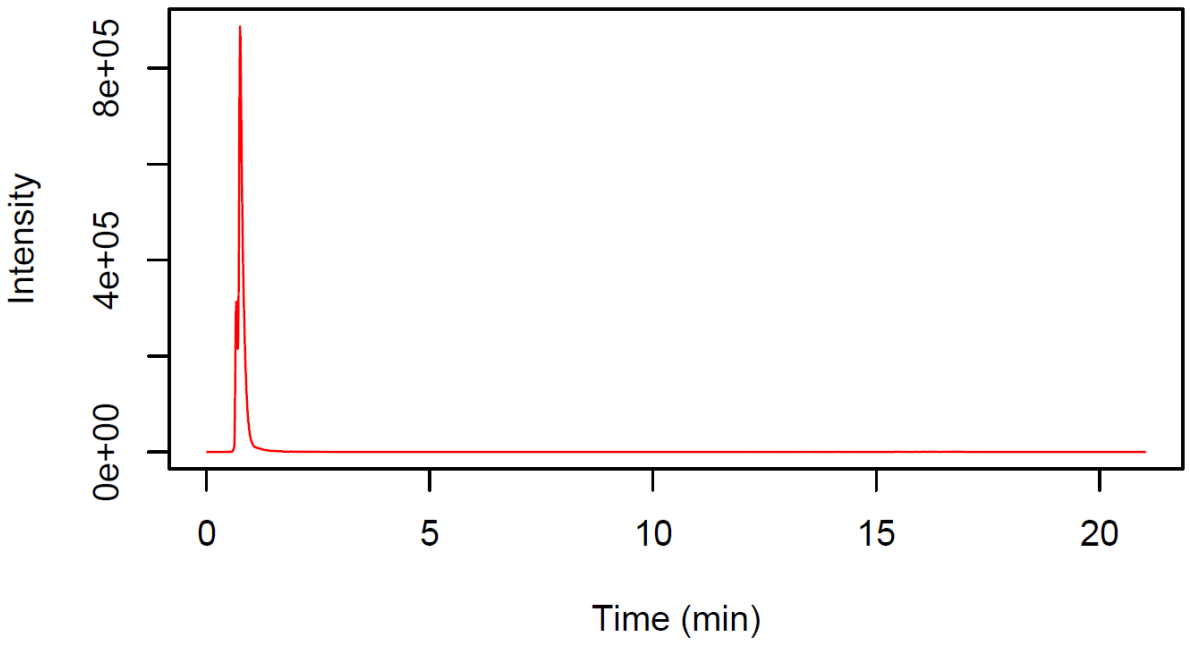


Figure A25: "Virtual" BPC traces illustrating co-condensation of G and H monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
156.0753	HGGH	H	[y1]1+	156.077	-0.002
156.0753	GHGH	H	[y1]1+	156.077	-0.002
156.0753	GGHH	H	[y1]1+	156.077	-0.002
195.0851	HGHG	HG	[b2]1+	195.088	-0.003
195.0851	HGGH	HG	[b2]1+	195.088	-0.003
195.0851	GHHG	GH	[b2]1+	195.088	-0.003
195.0851	GHGH	GH	[b2]1+	195.088	-0.003
213.095	HGHG	HG	[y2]1+	213.098	-0.003
213.095	HGGH	GH	[y2]1+	213.098	-0.003
213.095	GHHG	HG	[y2]1+	213.098	-0.003
213.095	GHGH	GH	[y2]1+	213.098	-0.003
252.1054	HGGH	HGG	[b3]1+	252.109	-0.004
252.1054	GHGH	GHG	[b3]1+	252.109	-0.004
252.1054	GGHH	GGH	[b3]1+	252.109	-0.004
270.1143	HHGG	HGG	[y3]1+	270.12	-0.006
270.1143	HGHG	GHG	[y3]1+	270.12	-0.006
270.1143	HGGH	GGH	[y3]1+	270.12	-0.006
293.1282	GGHH	HH	[y2]1+	293.136	-0.008
350.1525	GHHG	HHG	[y3]1+	350.157	-0.005
350.1525	GHGH	HGH	[y3]1+	350.157	-0.005
350.1525	GGHH	GHH	[y3]1+	350.157	-0.005

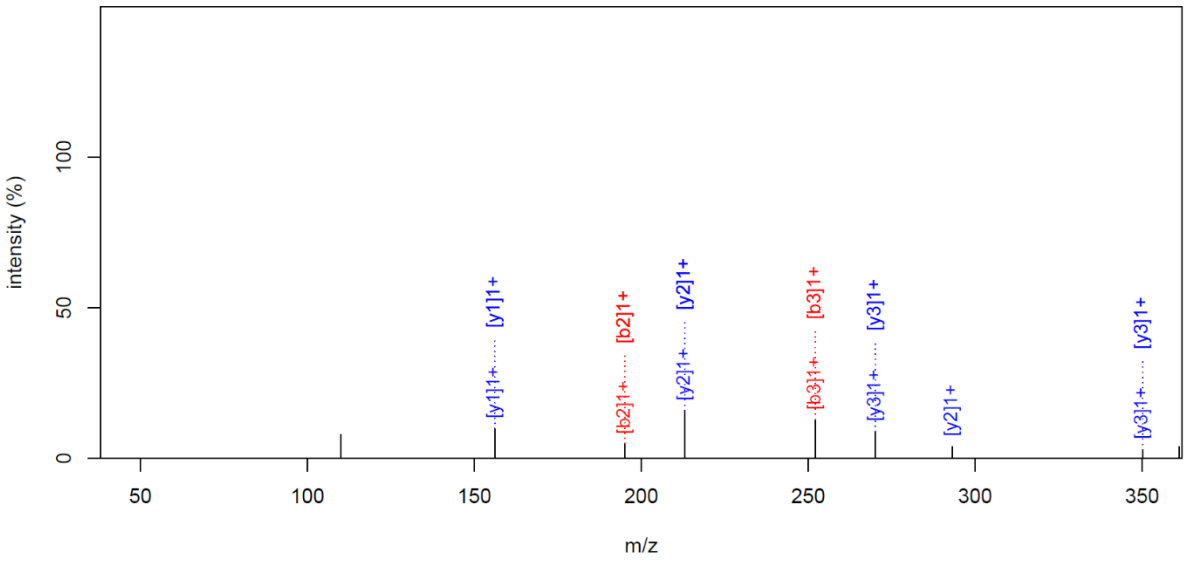


Figure A26: Example MS/MS analyses of products of G and H monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 407.179 (corresponding to a tetramer incorporating 2 x G and 2x H). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x H (ie. more than one sequence of the same composition is present, all incorporating G and H).

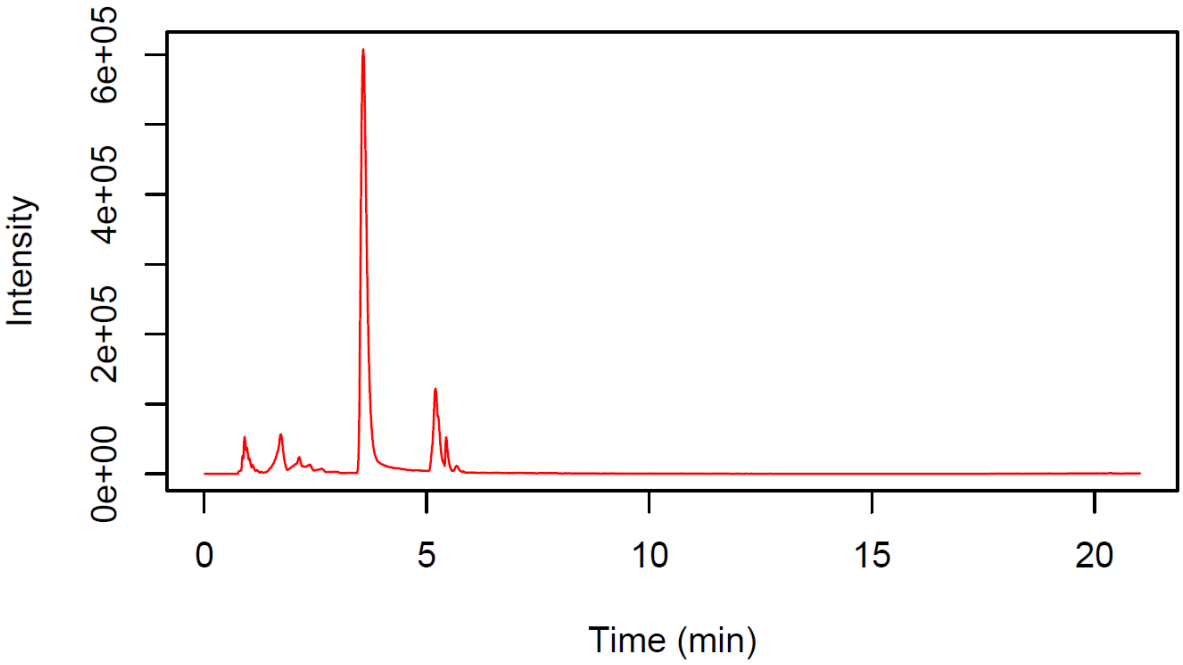


Figure A27: “Virtual” BPC traces illustrating co-condensation of G and P monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
116.0691	PGGP	P	[y1]1+	116.071	-0.002
116.0691	GPGP	P	[y1]1+	116.071	-0.002
116.0691	GGPP	P	[y1]1+	116.071	-0.002
155.0795	PGPG	PG	[b2]1+	155.082	-0.002
155.0795	PGGP	PG	[b2]1+	155.082	-0.002
155.0795	GPPG	GP	[b2]1+	155.082	-0.002
155.0795	GPGP	GP	[b2]1+	155.082	-0.002
173.0896	PGPG	PG	[y2]1+	173.092	-0.002
173.0896	PGGP	GP	[y2]1+	173.092	-0.002
173.0896	GPPG	PG	[y2]1+	173.092	-0.002
173.0896	GPGP	GP	[y2]1+	173.092	-0.002
212.1	PGGP	PGG	[b3]1+	212.103	-0.003
212.1	GPGP	GPG	[b3]1+	212.103	-0.003
212.1	GGPP	GGP	[b3]1+	212.103	-0.003
230.1104	PPGG	PGG	[y3]1+	230.114	-0.004
230.1104	PGPG	GPG	[y3]1+	230.114	-0.004
230.1104	PGGP	GGP	[y3]1+	230.114	-0.004

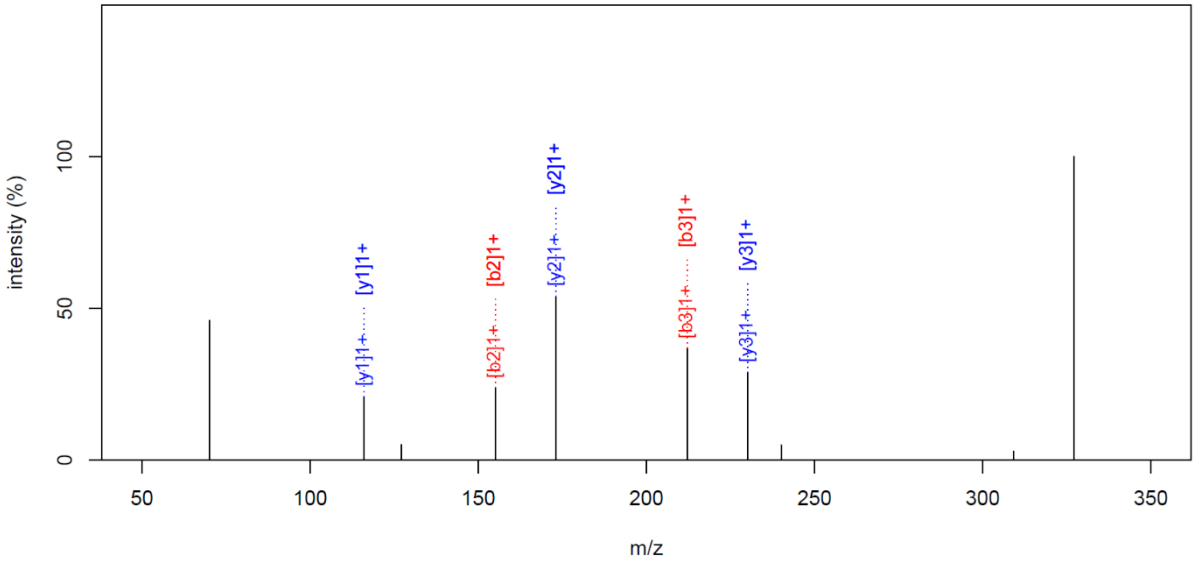


Figure A28: Example MS/MS analyses of products of G and P monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 327.166 (corresponding to a tetramer incorporating 2 x G and 2 x P). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x P (ie. more than one sequence of the same composition is present, all incorporating G and P).

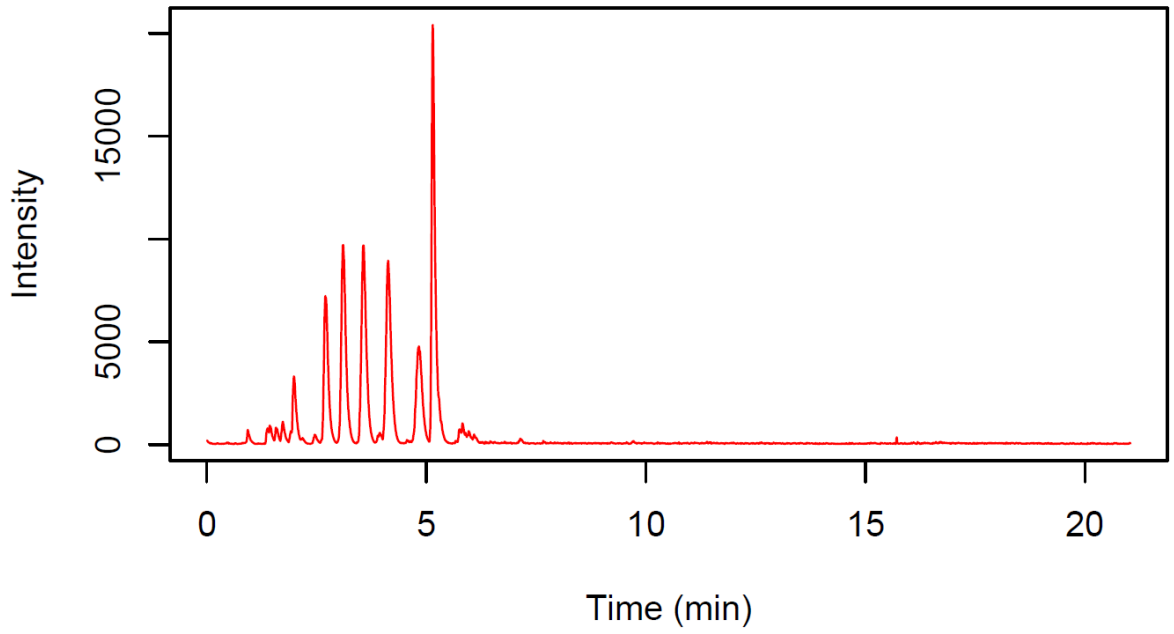


Figure A29: "Virtual" BPC traces illustrating co-condensation of G and V monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
115.0521	GGVV	GG	[b2]1+	115.05	0.002
118.0793	VGGV	V	[y1]1+	118.086	-0.007
118.0793	GVGV	V	[y1]1+	118.086	-0.007
118.0793	GGVV	V	[y1]1+	118.086	-0.007
157.0938	VGVG	VG	[b2]1+	157.097	-0.003
157.0938	VGGV	VG	[b2]1+	157.097	-0.003
157.0938	GVVG	GV	[b2]1+	157.097	-0.003
157.0938	GVGV	GV	[b2]1+	157.097	-0.003
175.1017	VGVG	VG	[y2]1+	175.108	-0.006
175.1017	VGGV	GV	[y2]1+	175.108	-0.006
175.1017	GVVG	VG	[y2]1+	175.108	-0.006
175.1017	GVGV	GV	[y2]1+	175.108	-0.006
214.119	VGGV	VGG	[b3]1+	214.119	0
214.119	GVGV	GVG	[b3]1+	214.119	0
214.119	GGVV	GGV	[b3]1+	214.119	0
232.1239	VVGG	VGG	[y3]1+	232.129	-0.005
232.1239	VGVG	GVG	[y3]1+	232.129	-0.005
232.1239	VGGV	GGV	[y3]1+	232.129	-0.005

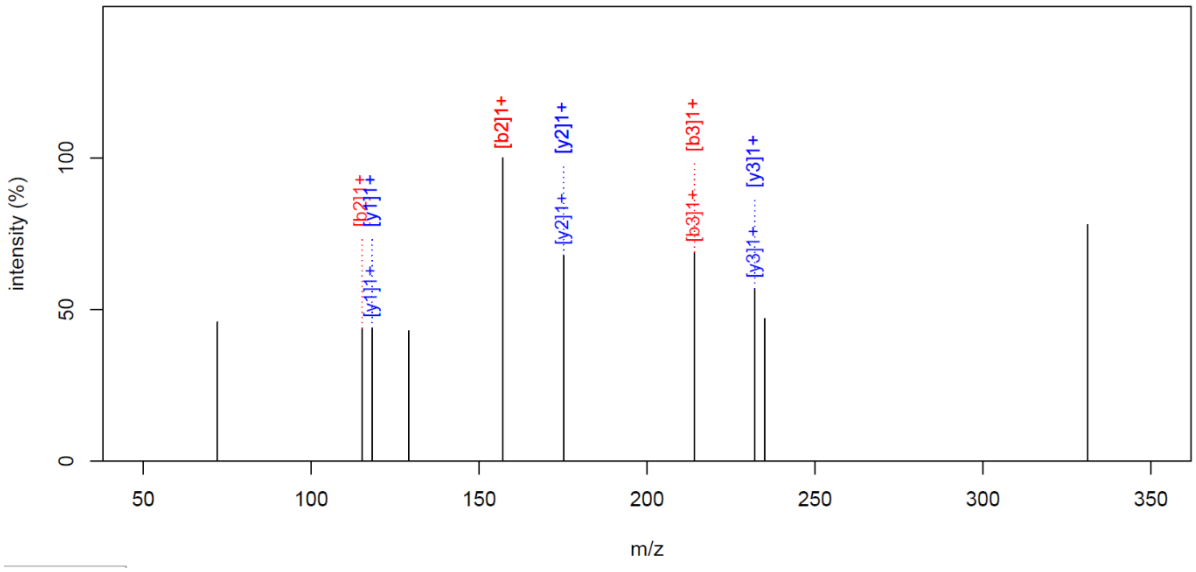


Figure A30: Example MS/MS analyses of products of G and V monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 331.198 (corresponding to a tetramer incorporating 2 x G and 2x V). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x V (ie. more than one sequence of the same composition is present, all incorporating G and V).

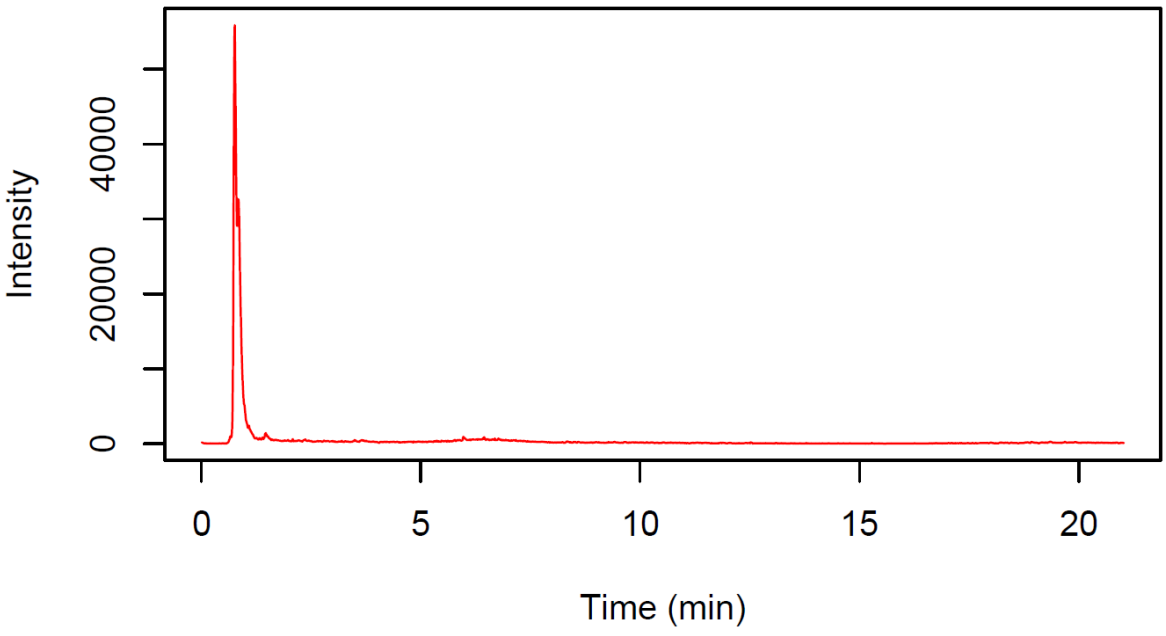


Figure A31: "Virtual" BPC traces illustrating co-condensation of G and T monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating both of the amino acids monomers.

Experimental Product ion m/z	Possible parent sequence(s) at m/z=363.115	Possible production sequence(s)	Fragment type	MSMS m/z	Error
102.055	TTGG	T	[b1]1+	102.055	0
102.055	TGTG	T	[b1]1+	102.055	0
102.055	TGGT	T	[b1]1+	102.055	0
115.0464	GGTT	GG	[b2]1+	115.05	-0.004
115.0853	GGTT	GG	[b2]1+	115.05	0.035
120.0686	TGGT	T	[y1]1+	120.066	0.003
120.0686	GTGT	T	[y1]1+	120.066	0.003
120.0686	GGTT	T	[y1]1+	120.066	0.003
133.0587	TTGG	GG	[y2]1+	133.061	-0.002
159.0758	TGTG	TG	[b2]1+	159.076	0
159.0758	TGGT	TG	[b2]1+	159.076	0
159.0758	GTTG	GT	[b2]1+	159.076	0
159.0758	GTGT	GT	[b2]1+	159.076	0
177.0841	TGTG	TG	[y2]1+	177.087	-0.003
177.0841	TGGT	GT	[y2]1+	177.087	-0.003
177.0841	GTTG	TG	[y2]1+	177.087	-0.003
177.0841	GTGT	GT	[y2]1+	177.087	-0.003
203.099	TTGG	TT	[b2]1+	203.103	-0.004
216.0929	TGGT	TGG	[b3]1+	216.098	-0.005
216.0929	GTGT	GTG	[b3]1+	216.098	-0.005
216.0929	GGTT	GGT	[b3]1+	216.098	-0.005
221.1111	GGTT	TT	[y2]1+	221.113	-0.002
234.1073	TTGG	TGG	[y3]1+	234.108	-0.001
234.1073	TGTG	GTG	[y3]1+	234.108	-0.001
234.1073	TGGT	GGT	[y3]1+	234.108	-0.001
260.131	TTGG	TTG	[b3]1+	260.124	0.007
260.131	TGTG	TGT	[b3]1+	260.124	0.007
260.131	GTTG	GTT	[b3]1+	260.124	0.007

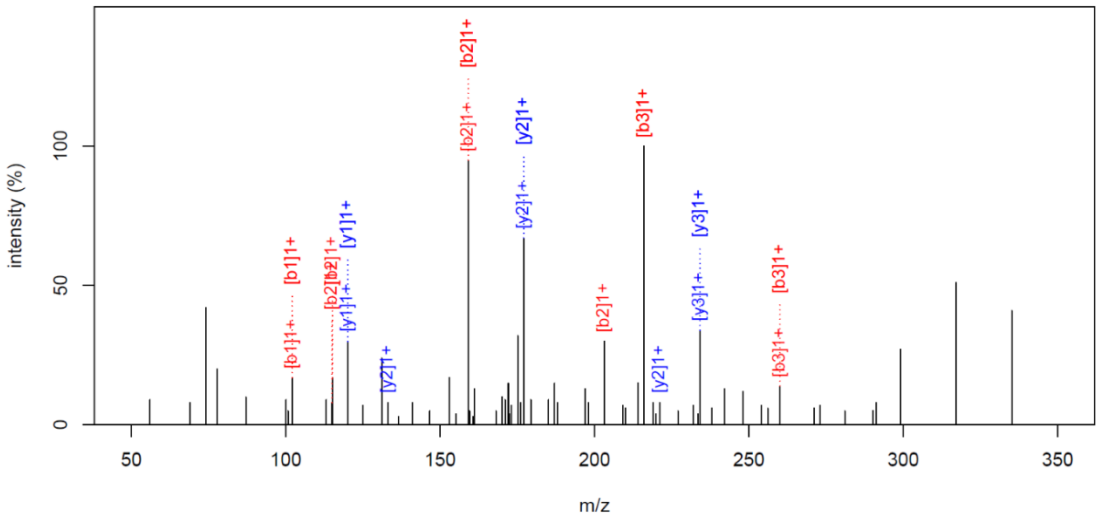
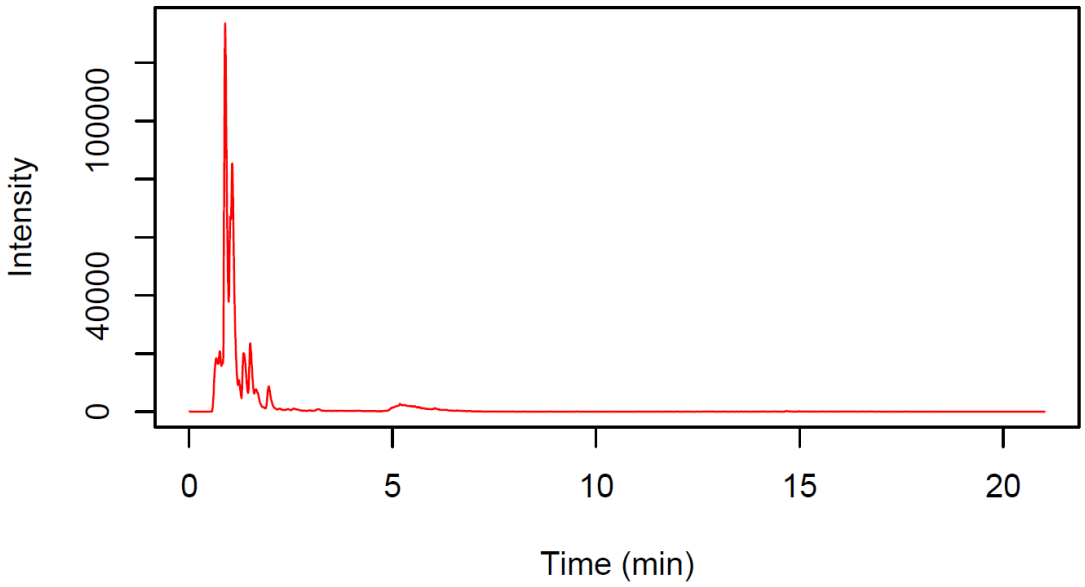


Figure A32: Example MS/MS analyses of products of G and T monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 335.156(corresponding to a tetramer incorporating 2 x G and 2 x T). A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G and 2 x T (ie. more than one sequence of the same composition is present, all incorporating G and T).

A)



B)

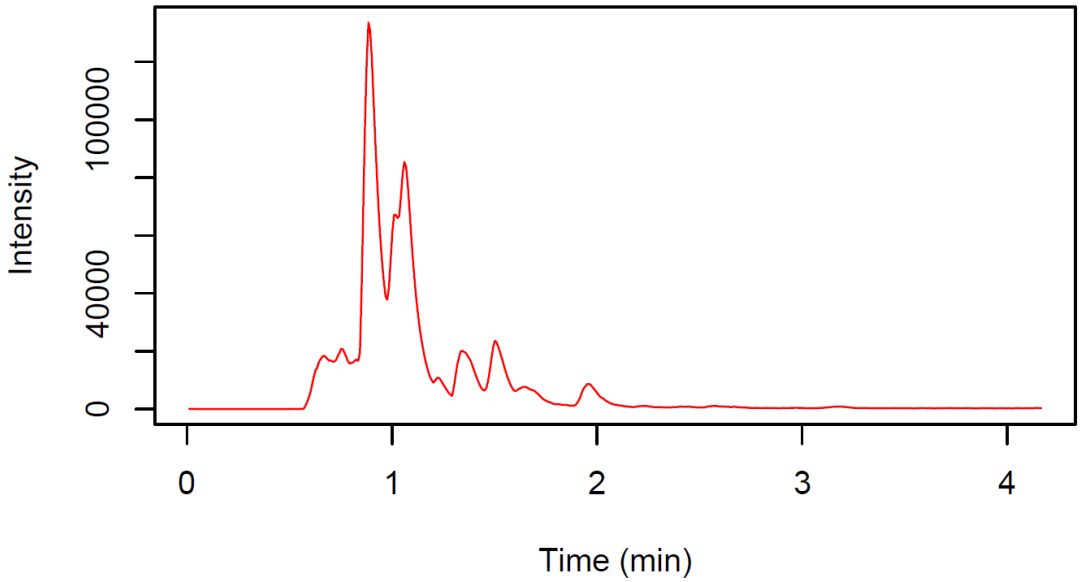
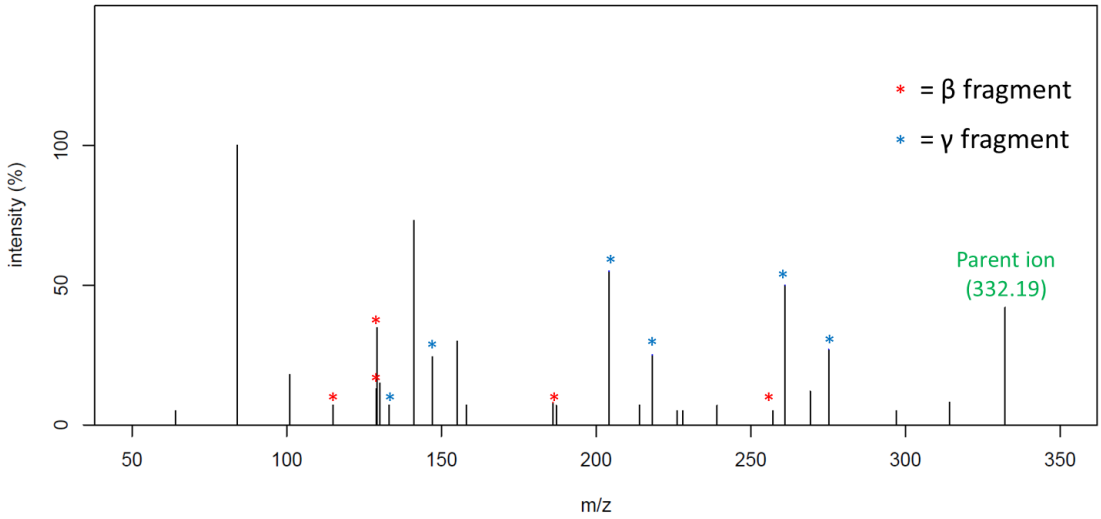


Figure A33: "Virtual" BPC traces illustrating co-condensation of G, A and K monomers. VBPC is constructed from the EICs observed for a list of putative oligomers incorporating all three of the amino acids monomers. A) shows the entire VBPC spectra and B) focuses on the region of the spectra where peaks from the co-condensation products are easily observed.



Experimental product ion *m/z*	Possible parent ion sequence(s) at *m/z* = 332.09	Possible product ion sequence(s)	Fragment type	MSMS *m/z*
115.0491	GGKA	GG	[b2]1+	115.05
115.0491	GGAK	GG	[b2]1+	115.05
129.0655	GAKG	GA	[b2]1+	129.066
129.0655	GAGK	GA	[b2]1+	129.066
129.0655	AGKG	AG	[b2]1+	129.066
129.0655	AGGK	AG	[b2]1+	129.066
129.1008	KGGA	K	[b1]1+	129.102
129.1008	KGAG	K	[b1]1+	129.102
129.1008	KAGG	K	[b1]1+	129.102
133.0597	KAGG	GG	[y2]1+	133.061
133.0597	AKGG	GG	[y2]1+	133.061
147.1115	KGGA	GA	[y2]1+	147.076
147.1115	KGAG	AG	[y2]1+	147.076
147.1115	GKGA	GA	[y2]1+	147.076
147.1115	GKAG	AG	[y2]1+	147.076
147.1115	GGAK	K	[y1]1+	147.113
147.1115	GAGK	K	[y1]1+	147.113
147.1115	AGGK	K	[y1]1+	147.113
186.0866	GGAK	GGA	[b3]1+	186.087
186.0866	GAGK	GAG	[b3]1+	186.087
186.0866	AGGK	AGG	[b3]1+	186.087
186.1216	KGGA	KG	[b2]1+	186.124
186.1216	KGAG	KG	[b2]1+	186.124

Experimental product ion *m/z*	Possible parent ion sequence(s) at *m/z* = 332.09	Possible product ion sequence(s)	Fragment type	MSMS *m/z*
186.1216	GKGA	GK	[b2]1+	186.124
186.1216	GKAG	GK	[b2]1+	186.124
204.1317	GAKG	KG	[y2]1+	204.134
204.1317	GAGK	GK	[y2]1+	204.134
204.1317	AGKG	KG	[y2]1+	204.134
204.1317	AGGK	GK	[y2]1+	204.134
218.1468	GGKA	KA	[y2]1+	218.15
218.1468	GGAK	AK	[y2]1+	218.15
257.159	KGAG	KGA	[b3]1+	257.161
257.159	KAGG	KAG	[b3]1+	257.161
257.159	GKAG	GKA	[b3]1+	257.161
257.159	GAKG	GAK	[b3]1+	257.161
257.159	AKGG	AKG	[b3]1+	257.161
257.159	AGKG	AGK	[b3]1+	257.161
261.1533	AKGG	KGG	[y3]1+	261.156
261.1533	AGKG	GKG	[y3]1+	261.156
261.1533	AGGK	GGK	[y3]1+	261.156
275.1684	GKGA	KGA	[y3]1+	275.171
275.1684	GKAG	KAG	[y3]1+	275.171
275.1684	GGKA	GKA	[y3]1+	275.171
275.1684	GGAK	GAK	[y3]1+	275.171
275.1684	GAKG	AKG	[y3]1+	275.171
275.1684	GAGK	AGK	[y3]1+	275.171

Figure A34: Example MS/MS analyses of products of G, A, and K monomer co-oligomerisation experiments. Spectrum (above) shows all the product ions resulting from MS/MS analysis of m/z = 332.09 (corresponding to a tetramer incorporating 2 x G, 1 x A, and 1 x K) eluted over t = 0.5 – 2 min in RP-HPLC-MS. A series of fragments are observed which can be assigned as those expected (below) from a number of different possible sequences which might be expected from tetramers incorporating 2 x G, 1 x A, and 1 x K (i.e. more than one sequence of the same composition is present, all incorporating G, A, and K).

7.1 Characterisation of different peptide populations:

[n.b. this process was created in collaboration with Dr. Andrew Surman, who wrote the R scripts]

To compare the different 'populations' of products produced by reactions in different environments, LCMS data was processed using the following procedure. Steps marked # were performed using a bespoke script written in R.

1. LCMS data acquired in Bruker proprietry format, as continuum data. Each experimental product (all experiments were performed in triplicate) was analysed in triplicate, leading a total of 9 data sets for each 'environment' tested.
2. LCMS data converted to 'open' 64bit .mzML format, using Proteowizard MSConvert. The instrument vendor 'pick peaks' option was selected to 'centroid' data (making files sizes manageable – ca. 450 Mb each).
3. [#] Peaks were identified and extracted using the Centwave function implemented in xcms[240]. This converts raw data to a series of 'features' or 'peaks', characterised by m/z, retention time (RT) and intensity. 'Permissive' settings were used to capture all possible features, including potential noise: low S/N (2), low noise gating (80 counts), high mzdiff (0.005 Da). 'Peak filling' was used, such that where a feature was observed in one sample, but not in the other samples to which it was to be compared, intensity data was extracted from the same m/z and RT.
4. [#] PC-DFA analysis was then performed on this feature data, to observe whether the populations were systematically the same or different (with scaling of intensity of individual features). This was represented in a 3D plot of individual data points, surrounded by a 'balloon' resulting from calculation of 95% confidence limit range.

In addition to these steps, the procedure was also repeated, with two extra steps:

3.b. [#] A list of all possible products resulting from combination of the amino acids included in each reaction (e.g. G + A + H) was constructed, and the m/z values expected for each calculated (using the OrgMassSpecR library).

3.c. [#] The features identified by the 'permissive' peak picking process were then filtered, removing all except those whose m/z values (+/- 0.05 Da) matched the combinatorial list of putative products.

In all cases, the PC-DFA plot resulting from the ‘unfiltered’ data resembled that resulting from filtering for putative ‘peptide’ product peaks in the sense that the same ‘populations’ were separated to similar degrees. The difference between the two approaches lay in the amount of ‘random’ variation unexplained by only a few principal components, reflecting the amount of ‘noise’ included in our unfiltered, permissively picked, data set. This demonstrates that the differences observed between the populations were a general phenomenon, not limited to a few ‘rogue’ species/artefacts; typically, filtering for putative peptide masses explained > 85 % of the inter-group variation.

See below a pictorial representation of the process:

a) Raw LC-MS data:

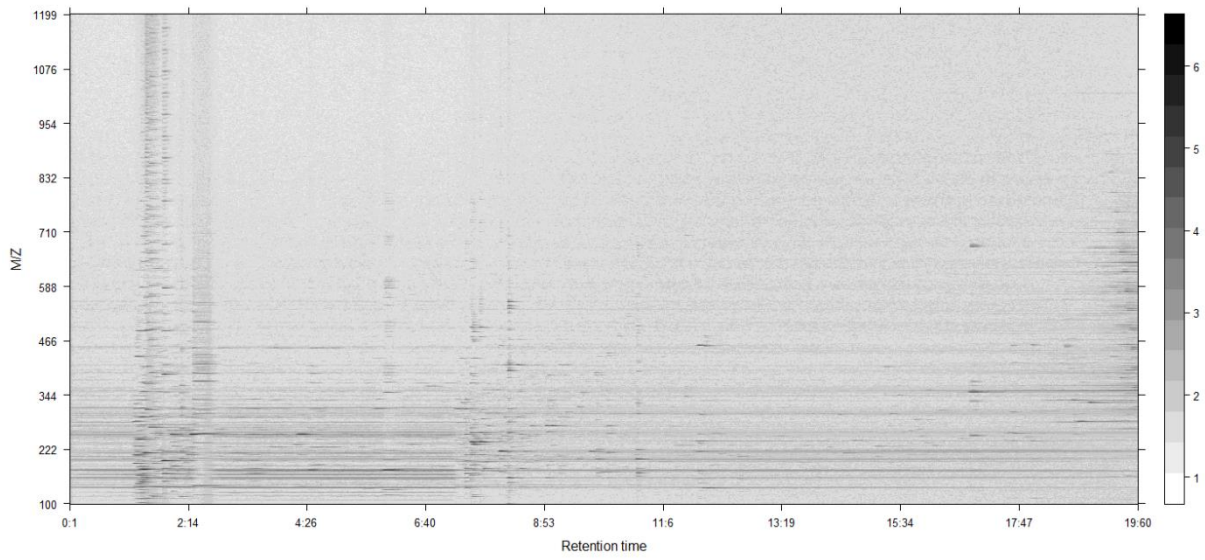


Figure A35: Raw LC-MS data. Full dataset contains many data points per scan (per 0.5 s; typically 8-20k).

b) Peak picking:

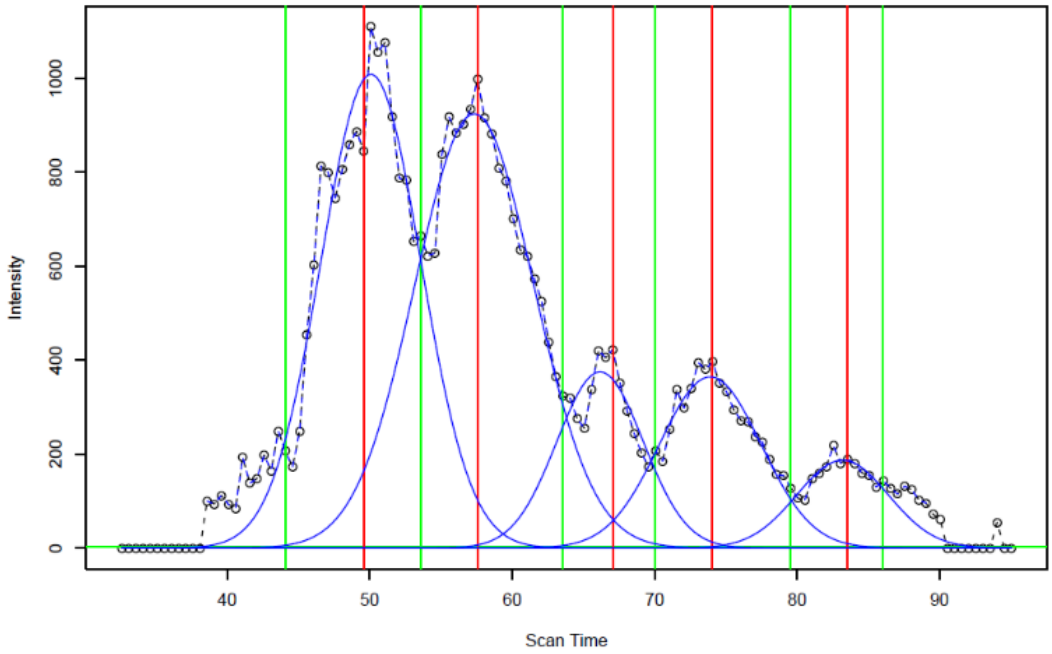


Figure A36: Peak picking representation.

Centwave algorithm essentially considers each possible EIC (every 0.005 Da) and looks for peaks which may be fit to Gaussian curves fitting selection criteria (S/N > 2; intensity > 80).

This produces a table of features:

						Intensities of different samples (Triplicate experimental replicates)					
mz	mzmin	mzmax	rt	rtmin	rtmax	APS.968a.1to10_7101_1722	APS.968b.1to10_7201_1723	APS.968c.1to10_7301_1724	APS.969a.1to10_7401_1725	APS.969b.1to10_7501_1726	APS.969c.1to10_7601_1727
190.0809	190.08	190.082	172.727	171.472	173.433	282.7021	305.1784	297.7177	451.0247	333.6133	279.2352
190.0811	190.08	190.0815	142.005	139.965	147.497	1732.586	1217.739	927.0057	1702.552	1766.161	1919.441
190.081	190.0804	190.0829	126.521	119.487	130.486	1082.344	3327.693	2309.189	1779.125	2167.089	2878.663
190.0891	190.0887	190.0906	100.047	87.053	105.065	23613.86	1895.013	1454.412	12396.86	12628.36	9743.705
204.0917	204.0913	204.0921	108.7695	101.532	113.998	31884.19	13816.25	8452.539	27572.9	23679.38	21882.96

Figure A37: Table of features for different experimental replicates.

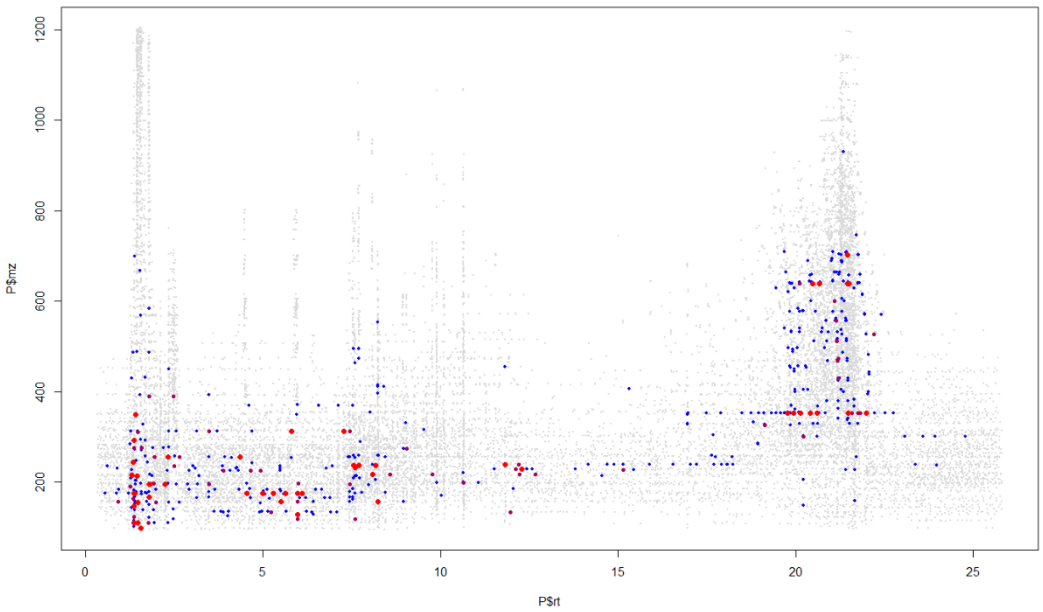


Figure A38: Plot of 'picked' features from the same LCMS dataset:
(low intensity features plotted in grey; progressively higher intensity features plotted in blue, burgundy and bright red)

Using relatively permissive picking criteria leads to the selection of a large number of features per LCMS dataset: typically 15-40k features.

c) Representative PCA results from 'unfiltered' and 'peptide mass filtered' data

	Percentage of variance	**Cumulative percentage of variance**			**Percentage of variance**	**Cumulative percentage of variance**
comp 1	61.27	61		comp 1	34.06	34
comp 2	18.17	79		comp 2	23.37	57
comp 3	5.82	**85**		comp 3	8.83	**66**
comp 4	1.97	87		comp 4	4.26	71
comp 5	1.45	89		comp 5	2.76	73
comp 6	1.29	90		comp 6	1.98	75
comp 7	0.84	91		comp 7	1.25	77
comp 8	0.72	92		comp 8	1.03	78
comp 9	0.63	92		comp 9	0.92	78
comp 10	0.52	93		comp 10	0.86	79

Figure A39: PCA results from data from mineral reactions (unfiltered – left, 21666 features picked; peptide mass filtered– 798 features

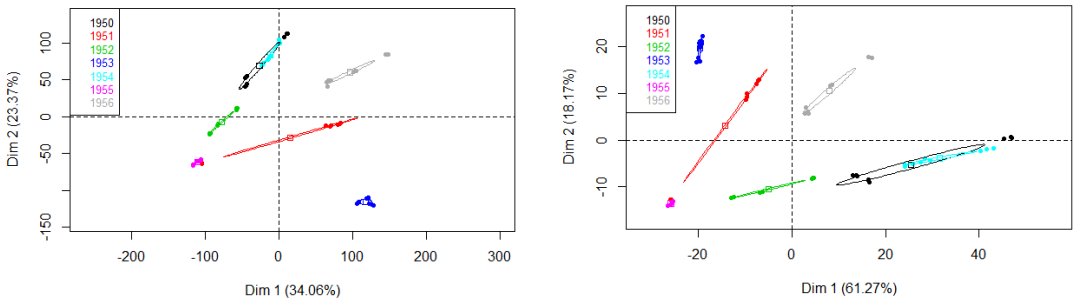


Figure A40: PCA 2D representation from data from mineral reactions (unfiltered – left, 21666 features picked; peptide mass filtered– 798 features.

While different amounts of variation are explained by PCA of the filtered and unfiltered feature lists, the separation between the groups yielded is qualitatively similar.

PC-DFA sharpens the resolution between individual groups, while leaving separation qualitatively similar to simple PCA.

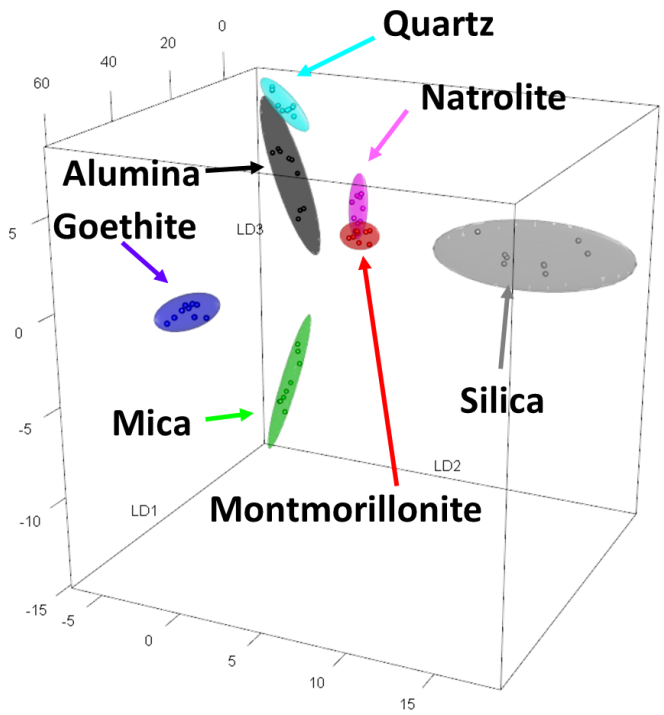


Figure A41: PC-DFA representation from data from mineral reactions (peptide mass filtered– 798 features).